DOCUMENT RESUME

ED 172 656 **HE 011 393**

AUTHOR	Andresen, L. W., Ed.; Boud, D. J., Ed.
TITLE	Aspects of Student Assessment. Proceedings of a Conference Held at the University of New South Wales, 18-20 July 1978. Tertiary Education Research Centre Occasional Publication No. 13.
INSTITUTION	New South Wales Univ., Kensington (Australia). Tertiary Education Research Centre.
PUB DATE	Dec 78
NOTE	113p.; Not available in paper copy due to marginal legibility of original document
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Course Organization; Criterion Referenced Tests; *Educational Assessment; Essays; Evaluation Criteria; Evaluation Methods; Feedback; *Grading; Higher Education; Laboratory Training; Nongraded Classes; Plagiarism; Planning; Preservice Education; *Self Evaluation; *Student Evaluation; Student Participation; Test Results

ABSTRACT

A conference was held by the University of New South Wales, Australia, to enable staff and students to examine the purposes of student assessment, implications of various forms of assessment, and ways in which assessment might be most effectively conducted. Eric Daniels spoke on the jury system of assessment in architecture, Raja Bandaranayake spoke on the need for a rational and analytical approach to the planning of assessment, and Dixie Coulton spoke on student participation in the assessment process. Study groups considered the following areas: grading/assessment, continuous assessment, assessment as feedback to students and staff, student participation in assessment, and assessment of class participation. Workshop reports presented are: "Diagnostic and Other Feedback Based on Exam Analysis," by Max Dunstan; "Assessment of Laboratory Work," by Ian Sefton; "Assessment in Pre-Service Teacher Education Programmes," by Lee Owens and David Smith; "Essay Marking," by Stan Ross; "How to Organise a Course without Grading," by Dave Boud; "Selecting Appropriate Assessment Methods for a Course," by Jackie Lublin; and "Criterion-Referenced Assessment," by Michael Bailey. A bibliography, rules of the law school on the use of the writings of others and plagiarism, and a paper on teaching the writing of academic papers by Peter Bamford are appended. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED172656

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Tertiary Education Research Centre of New South Wales

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ASPECTS OF STUDENT ASSESSMENT

Proceedings of a Conference held at The University of New South Wales 18–20 July 1978

Edited by L.W. Andresen and D.J. Boud

NE 011393

ISSN 0313-5810

Occasional Publication No. 13
Tertiary Education Research Centre University of New South Wales

ASPECTS OF STUDENT ASSESSMENT

Proceedings of a Conference held at the University of New South Wales

Kensington

18-20 July 1978

Edited by
L. W. Andresen &
D. J. Boud

Occasional Publication No. 13
Tertiary Education Research Centre
University of New South Wales
December 1978

Published by the Tertiary Education Research Centre
University of New South Wales
Kensington

© TERC December 1978

National Library of Australia Card No. &
ISSN 0313 - 5810

CONTENTS

	PAGE
PREFACE	(i)
SECTION I : PAPERS	
Chairperson's remarks : Professor N.L. Svensson	3
Assessment of Architectural Design in the Faculty of Architecture : Professor E.C. Daniels	4
A Rational Approach to the Assessment of Student Performance : Dr. Raja Bandaranayake	9
Some Perspectives on Assessment : The Student Point of View; Ms. Dixie Coulton	15
SECTION II : PANEL DISCUSSION	
Response by Discussant : Dr. John Kleinig	21
Questions from the Floor	26
SECTION III : STUDY GROUP REPORTS	
Study Group Participants	41
Grading ≠ Assessment	42
Continuous Assessment	45
Assessment as Feedback to Students and Staff	48
Involving Students in Assessment Procedures	50
Assessment of Class Participation	52

CONTENTS

PAGE

SECTION IV : WORKSHOP REPORTS

Diagnostic and other Feedback based on Exam Analysis: Dr. Max Dunstan 57

Assessment of Laboratory Work : Dr. Ian Sefton 59

Assessment in Pre-Service Teacher Education Programmes: Dr. Lee Owens and Mr. David Smith. 62

Essay Marking : Mr. Stan Ross 65

How to Organise a Course without Grading: Dr. Dave Boud 69

Selecting Appropriate Assessment Methods for a Course: Ms. Jackie Lublin 72

Criterion- Referenced Assessment : Mr. Michael Bailey 74

SECTION V : APPENDIXES

Some Notes on Criterion-Referenced Measurement: Mr. Michael Bailey 79

Teaching the Writing of Academic Papers : A Contribution to Discussion : Mr. Peter Bamford 93

The Problem of Plagiarism: U.N.S.W. Law School 98

Bibliography: Aspects of Student Assessment 102

List of Conference Participants 107

PREFACE

The Tertiary Education Research Centre mounts an annual conference choosing for its theme some topic of immediate concern to university staff and students. The theme of the conference for 1978 was Aspects of Student Assessment. The aim of the conference was to bring together staff and students from all parts of the university and elsewhere to examine and discuss the purposes of assessment, the implications of various forms of assessment and the ways in which assessment might be most effectively conducted.

The conference extended over three days and was planned in three major sections corresponding to these aims:

i. Introduction to the theme:

Professor Noel Svensson, Chairman of the Faculty of Engineering, U.N.S.W. chaired the opening session. First Professor Eric Daniels, Head of the Department of Undergraduate Studies in the School of Architecture spoke on the jury system of assessment in architecture. Subsequently, Dr. Raja Bandaranayake of the Centre for Medical Education Research and Development spoke on the need for a rational and analytical approach to the planning of assessment, and Ms. Dixie Coulton, President of the U.N.S.W. Students' Union spoke on student participation in the assessment process. Dr. John Kleinig of the School of History, Philosophy and Politics at Macquarie University acted as discussant to these papers and introduced a discussion involving the panel and participants.

ii. Study Groups:

In the afternoon of the first day a series of study groups led by T.E.R.C. staff were offered as an opportunity for participants to explore certain issues in greater depth. These comprised

Grading ≠ Assessment
Assessment of Class Participation
Continuous Assessment
Assessment as Feedback to Students and Staff
Involving Students in Assessment Procedures

iii Workshops:

The remainder of the conference consisted of two days of workshops activities. Participants attended workshops devoted to areas of assessment in which to develop their skills. These were led by people with special experience or expertise in each area. In addition there was an opportunity for participants to initiate workshops on topics of their choice and, in response to this, a workshop on Assessment in Teacher Education was convened by two staff from the School of Education at the University of Sydney. The other workshops were:

Essay Marking
Assessment of Laboratory Work
Diagnostic and other Feedback based on Exam Analysis
Selecting Appropriate Methods of Assessment for a Course
Criterion-referenced Assessment
How to Organise a Course Without Grading

This volume represents the proceedings of the conference. For the most part it contains accounts by participants, of the various study groups and workshops in which they took part. It also contains the papers presented on the first day plus supplementary material which was tabled in workshops and a list of useful references.
In preparing for the conference the organising committee became aware of the paucity of published Australian material on assessment in tertiary education. We hope that this publication might assist those who wish to explore some of the current issues in this area.

Dave Boud
Lee Andresen
July 1978

SECTION I

PAPERS

CHAIRPERSON'S REMARKS:

Professor N.L. Svensson, School of Mechanical & Industrial Engineering
University of New South Wales.

The matter of assessment is now a very important one. I recollect a booklet I read some years ago on the aims and importance of examinations (or what we now call assessment) and their relationship to the nature of education. I will simply quote from that book by Professor Gasking

> 'Broadly speaking an educated person is more than an uneducated one (A) the *master* of certain *skills* (B) *'cultured'* and (C) the possessor of a *'trained mind'*.

As far as a trained mind is concerned, this can be broken down further. He is able to

> 'find things out for himself',
> 'organise his knowledge and ideas systematically'
> 'generalise safely', and 'reason'. *

The importance of these is that there is a fairly strong relationship between the way one examines or assesses and the way one teaches or learns. One cannot aim to teach people to become educated or cultured or to have certain skills or a trained mind, without expecting that what we do is going to be reflected in the way students approach the subject. Their concern, very significantly, is with being rewarded for their studies and this reward comes through our processes of assessment.

So assessment is a very significant part of the whole educational scene. We are going to hear several points of view on it, after which we should have some indication of the width and importance of the subject.

---

*Gasking, D.A.T. Examinations and the aims of education, Melbourne 1948

THE ASSESSMENT OF ARCHITECTURAL DESIGN IN THE FACULTY OF ARCHITECTURE

by Professor Eric C. Daniels, School of Architecture, University of N.S.W.

The vast majority of students enrolling in courses in architecture intend to qualify as architects and enter the profession. The teaching of subjects in the School involves two main aspects - the scholastic rigour of the subject in its own right, and its relevance to the practice of architecture. There is emphasis placed upon small-group teaching , discussion, processes, analysis, critical evaluation and problem solving.. Marking is played down but assessment (as a grading) is considered important because it reflects a judgment of the value of the work produced. All subjects are continuously assessed and some subjects have final examinations.

A number of methods of testing are used in the various subjects of the course and depend upon what the lecturer seeks to examine - knowledge of facts, understanding of principles and methods, skill to communicate, and the ability to analyse and synthesise.

I wish to deal with the assessment procedures that are used for the subject of architectural design. Design is the subject which is the amalgam of other subjects studied. The subject involves analysis and the synthesis of a large number of interrelationships together with degrees of creativity. Each student believes his solution to be original and certainly no two solutions are the same, although the range of solutions generally fall into a small number of types. It may well be that one of the 'types' of solution may be 'best' and it may well be that the highest assessments are given for designs other than those of the 'best type' - simply because, on the balance of the factors and criteria, the degree of resolution is greater in the scheme presented than is evident from the work that any students have produced whose design is of the 'best type'.

What is considered important in the assessment of Design is that the student's work receives proper criticism - the grading given is

a reflection of the value of the work done. Since each student's work is original, he reacts more positively to criticism of his work much more than to criticism of someone else's work that he may or may not recognise as having the same faults or good points as his own.

In the past the assessment of Design solutions was made in camera by a panel of lecturers or a single lecturer. Two weeks or more following the submission of the designs, a 'public' criticism would be given of a range of the schemes presented - such crits could be of 1 to 2 hours duration; an exhausting session for lecturers as well as students, and often so late that students had 'forgotten' the project.

The desirable features that are considered necessary for Design assessment are (1) schemes should be able to be ' presented' by the student to the assessors; (2) the 'presentation' should occur as soon as possible after the scheme is submitted; (3) the student should have the opportunity to answer any criticism; (4) there should be more than one critic and assessor; (5) the assessment should be made known to students within one week of the submission of their schemes.

The method for assessment of Design that has been developed in the School is that all major and most minor design projects are submitted to a jury.

The process is : a schedule is prepared that allocates a time, place and duration for each student; the student presents his scheme; questions are asked; criticism is offered by the jury; discussion occurs between designer and jury; the jury makes a graded assessment; the lecturer averages these assessments; the grades are published in the studio.

The time allocated for the consideration of each scheme varies with the degrees of complexity of the problem - usually it is between twenty minutes and one hour.

The composition of the jury varies with the level of the student in his course and the nature of the project. Usually a jury is three, the student's lecturer and two other staff members. Juries for junior students generally comprise 'sympathetic' jurors, i.e. those who are very familiar with the project and the process the student has gone through to evolve the solution. Senior students usually have a hostile jury - the student's lecturer, a faculty staff member who has not been involved in the process of the solution, and a practising architect or specialist in the building type. Often, and more especially for minor projects, a student becomes a jury member. Some effort is made to brief jurors on the problem before they face the student's presentation.

The system is a good one from all points of view but has a number of pressures upon it that are not always easily overcome. The whole process is very time consuming when it is understood that any one year of students is about 100 - and there are four years. Efforts are made to spread the times of the Schedule but the end of Session is a hectic time. The time allocated for each student's scheme is usually shorter than is needed - and it needs a firm jury leader to keep to the Schedule. The briefing of jurors needs care if the jury session is to be productive.

In spite of these difficulties, the method is one that I believe to be worth the effort. Jury sessions are the occasions where students may learn a great deal about architecture, where philosophies and attitudes may be aired and tested in the context of a particular design scheme, and where students develop a responsibility for the work they produce.

QUESTIONS FROM THE FLOOR

Question:

About the time involved; presumably that subject is one of a number of subjects; and if you were speaking of the school or faculty as a whole that would involve some five hundred students over five years. Is it possible to give us some idea of the time input for assessment of all types to take place in the school, against the amount of time put in on teaching? Some ratio between the two? And, if there are various modes of assessment used, whether this particular form takes more or less time than the others?

Professor Daniels:

The system I have just talked about must take considerably longer than where a subject may have just a series of quizzes. It is continuous assessment all the way through. A one-session subject may have, say, three quizzes and finish with a final examination. It would use about two or three hours a week teaching time and each quiz would take about two hours with the final exam of three hours. That's nine hours of assessment out of fifty two hours teaching, or five hours teaching to every one assessment. Now in Architectural Design I would say we do use a little more time than that. It would depend on the subject but the method would not require as much time as one would think at first.

Question:

Do those who get an unsatisfactory jury judgement have a chance to go back to the drawing board and resubmit?

Professor Daniels:

It varies with lecturers and how they run their particular subject. It is generally true of junior and of senior students. Juniors

can be given the chance to correct and put into effect the jury's criticism on their work but then the student seldom gets a grade better than pass regardless of the quality of the final work. With seniors the same opportunity may be given because it is near graduation time so we must compare the value of giving them the opportunity to bring it up to the required standard. Those opportunities are not generally used for a reason associated with the question that was asked earlier; how much time can you afford to spend on it? We would have no objection to students doing as you have suggested but we would rather like to be doing some other things than just assessing and criticising Architectural Design.

Question:

From our legal system the word jury embodies the idea of the reaction of others to an argument which, in the most serious cases, could really be a matter of life and death. I wonder if, when we submit creative work often enough to the reactions of others who tell us 'This will do' or 'This will not do' then do we not encourage in our students an over-trust in the reactions of experts outside themselves and a distrust of their own hunches, intuitions and innate originality? If repeated often enough is this more a disservice than a service to these students?

Professor Daniels:

I think I understand fairly well what you say. If a talented student comes to the School of Architecture we hope we won't destroy that talent. I think the School is set up essentially for those who are adequately intelligent to ultimately be able to carry on a professional life and I think the whole organisation of the School is to try to bring such persons to professional competence. We have, fortunately, a few such highly tuned people as you refer to. In a good deal of small group teaching one gets to

know students fairly well. One of the great 'arts' is, I think, to be able to find in Schools of Architecture good critics of architecture, people able to criticise work without criticising the person who produced it. I think the chances of what you say happening are indeed there but not with the seriousness of outcome you have suggested. There is the possibility of it for those very sensitive students. As a professional person you are there as a servant of the community. You react to them. So whatever creative work you do needs to have the concurrence of somebody else. This might be comparable to the jury. That jury may or may not be expert. They may simply be members of the public. The jury is a fact of life in practicing Architecture. Yes, there is a possibility of a student being upset by criticism. That's always the case, whatever the system. Because of its formality the jury system may increase that possibility but there is a general understanding of these things by jurors and by students. Yet at the same time the student needs to understand that's what he's up against in life anyway.

A RATIONAL APPROACH TO THE ASSESSMENT OF STUDENT PERFORMANCE

by Dr. Raja Bandaranayake, University of New South Wales

It is my purpose to present to you a rational approach to the assessment of student performance, as distinct from a problem-centred approach. The latter approach is particularly useful when a gap is perceived between an observed and ideal state of affairs, as for example when our experience in using an existing system of examination shows us that something is wrong with it. The rational approach is particularly useful when we plan something new, as for example when we plan an examination for a group of students. This presentation is confined to the rational approach, and I will be using examples from medical education, because this is the area I am most familiar with.

Let us take each step in the rational approach and see how we can apply it to the assessment of student performance, and what factors have to be taken into consideration in such application. As I go along, I will raise certain issues and questions concerning common practices at present and, hopefully, some of these issues would lend themselves to discussion later on in this conference.

As in planning any venture on a rational basis, the first step is to identify the purpose of assessment. In a general sense we may say that it is to find out if the students are achieving/have achieved the objectives of the course. But why do we want to find this out? This is an important question to ask because the way you set about making the assessment is to an extent determined by the answer to this question. We assess student performance for different reasons e.g. for selection, for certification of some professional capability, for determining readiness to proceed to the next step in the sequence of the program, for comparing students with one another (i.e. to rank-order or categorize them), for comparing two groups of students who have been exposed to say, different methods of instruction, for diagnosing deficiencies in student learning with the intention of correcting those deficiencies, for obtaining information as to the effectiveness of our teaching, and so on. It is important to be clear in our minds what is the purpose of the particular assessment that is planned. In a general sense the above purposes fall into two broad categories:

1. those which are concerned with determining how students are performing as learning progresses, i.e. progressive or formative assessment;

2. those which are concerned with determing how students who have been subjected to a program perform at the termination of that program i.e. terminal or summative assessment.

I observe with interest that one study group which has been proposed for this afternoon is on 'Continuous Assessment'. A dilemma which I have always been faced with has to do with the purpose of continuous

assessment. The purpose of such assessment is to obtain feedback both for the student and for the teacher, so that corrective measures could be taken in time. The more informal such assessment is the better, since the element of threat that is inherent in any assessment is minimized, and students see it as part of their learning rather than of examination. It has been my experience, however, (and perhaps there are cultural differences here) that students often do not take such informal assessment seriously, and hence the validity of the feedback one obtains is questionable. On the other hand, if such assessment is formalized (as for example when the results of such assessment contribute to the making of critical pass/fail decisions), the element of threat is so exaggerated that feedback is again questionable. This is perhaps an issue which could profitably be discussed this afternoon.

Another issue which needs consideration is whether, in fact, there is a need for summative assessment. It is my belief that there is. Even if mastery of sub-units within a unit of instruction can be assessed formatively as the course progresses, there is a need to determine whether a student has been able to relate these sub-units where such inter-relationships exist, and this can only be done at the termination of that course. In medical education for example, where each sub-unit could be an organ system, mastery of each does not necessarily mean that the medical student is capable of understanding the relationships between these organ systems, and this can only be assessed terminally.

The next step in the rational approach to assessment is to determine what it is you want to measure. Learning consists of three components - knowledge, skills and attitudes. We value all three in education. The relative emphasis we place on each component would vary from subject to subject. For example, the assessment of attitudes is more likely to be important in the clinical subjects than in the basic medical sciences. One caution though - in identifying what it is we want to measure we should not confuse the objectives of learning with the methods of achieving those objectives. For example, if we want our students to learn about

the anatomy of the shoulder joint, and one way of achieving that is to get them to dissect the shoulder joint, then what we want to measure is not their ability to dissect the shoulder joint but their level of knowledge of the anatomy of the shoulder joint.

What do we exactly mean when we say that we want our students to know something? Do we want them to be able to recognize it, understand a concept, apply facts pertaining to it to a novel situation, or to use these concepts and facts to solve a problem? This definition of what we want to measure in operational terms is the next step. Unless we identify which of these we want to measure we will not be sure whether our tests are in fact measuring what we want them to measure. If, for example, we want third year medical students to recognize a cirrhotic liver, we will not ask them to describe one, but see whether they can identify a pathological specimen.

This definition in operational terms is prerequisite to the next step, which is to devise the appropriate test to make the measurement. We must ensure that the test we select is measuring exactly those abilities in the student that we want it to measure, i.e. it is a valid test. Getting the students to describe the appearance of a cirrhotic liver is invalid for measuring the ability 'being able to recognize a cirrhotic liver'.

If a particular ability can be measured in several ways, then you have to decide which is the most reliable way, i. e. which method is least likely to be affected by extraneous factors which affect the measurement such as examiner bias, student anxiety or tiredness etc. In other words, which is the method that is most likely to give reproducible results Suppose, for example, that you want to find out whether the student can recall the antibiotic of choice in the treatment of typhoid fever. This could be determined in an oral examination, as part of an essay question or as an MCQ. All three are valid tests to measure this ability, but the oral is least likely to give reproducible results, the essay less so, and the MCQ most so. Remember, you have a choice only after you have determined

which tests are valid. In other words, reliability has to take second place to validity.

We have heard so much about the virtues of MCQ, often as though they were the panacea for student assessment. There are in fact some situations where the only test of competence used in medical education is the MCQ. It should be clearly understood that MCQs do have only a limited place in student assessment, but when they do they measure very reliably. But then, again, reliability has to take second place to validity.

Let's say we want to determine how clinically competent a final year medical student is. This composite ability called clinical skills has several components.

1. To recall certain facts and concepts
2. To apply them in a novel situation, such as with a patient
3. To take a medical history
4. To do a physical examination
5. To be sensitive to the patient's needs and feelings
6. To know what investigations to call for
7. To carry out certain simple ward tests
8. To interpret laboratory data
9. To use the data at his disposal to make a diagnosis
10. To prescribe the appropriate regime of management.

Components 1, 2, 6, 8, 9 & 10 can be tested with a pen-and-paper test. Of these some lend themselves to an MCQ test. 3 & 4 can only be tested by presenting a patient, and 7 by getting the student to perform the appropriate test (s) . These require observation by the examiner of the student's actual performance. Component 5 can be tested by observing the student, preferably over a long period of time.

It is also best to select a method which measures what you want to measure as directly as possible.

Once you have decided on the method of testing you have to collect the data. One question that has to be answered is 'who should collect the data?' How often do we use other students - both peers and senior students - for this purpose? The new medical school at Newcastle uses peers for assessment, with certain guidelines being provided. How often do we let the student himself do a self-assessment? Very often answer scripts are treated with so much confidentiality that students don't even know where they have made errors. There is in fact, something called the ' moment of learning', and this is generally just after an examination, when students are most keen to know 'how they have fared'. What is learned then is often retained well

You now have to set standards of acceptable performance. A decision has to be made whether you set absolute criteria against which you judge each student's performance or relative criteria, as when judging each student's performance in relation to the performance of the student group as a whole. If the former, criteria should be pre-specified, preferably in consultation with your colleagues.

Next, summarize and report the results giving particular consideration to the recipient of the report and the form in which the report is made.

The final Step is to make judgement on the measurement, in relation to the purpose identified in Step 1. For example, if the purpose of the measurement was to provide feedback to students as to their strengths and weaknesses, it is virtually useless to make a judgement that a particular sutdent has passed or failed your test. Rather, you should indicate the content areas he was strong in, and those he was weak in.

SOME PERSPECTIVES ON ASSESSMENT: THE STUDENT POINT OF VIEW

by Dixie Coulton, President, U.N.S.W. Students' Union.

First of all let me say that in my opinion there is too much formal assessment in this University. The outside world contains very little of this method, and perhaps we should begin our discussion today by trying to ascertain whether or not formal assessment fits into the modern world, and is as relevant as it used to be. I would maintain that it does not and is not.

What we feel is required today is a great deal more of in-depth discussion and critical analysis of what is taught. Perhaps we could also consider whether or not we are attempting to cram too much material, too quickly, too superficially, into the heads of most of our students. I believe that we are.

More attention should be paid to the relevance of the material itself. If it can not be shown to be meaningful to our present society, we feel that students tend to lose interest. Analysis of the 1978 alternative handbook yielded numerous complaints from students in a wide number of disciplines that they did not know why they were being taught particular material in a particular way. Now whether that was because it had not been properly explained to them, or it was indeed the case that the material was not relevant, is not entirely clear but the point that I wish to make clear is that they were not sufficiently involved in the assessment and learning process, and thus were rendered incapable of forming a proper judgement. It is this absence of participation that causes students to lose interest in what they are learning.

In summary, I feel that there is too much formal, irrelevant assessment. Students should be at a University for much more than just a piece of paper. There should be room in their timetables, and I mean that it should be consciously set aside, for other, wider interests.

It could be argued that these other areas of interest, for example involvement in clubs and societies, student union politics, sport or even just arguing in the refectory, all serve to help prepare students for a more active, healthy involvement in society. Assessment procedures should be structured to recognise this fact.

As most of you are well aware the nature and structure of the workforce is at present undergoing rapid change. Employers have used this as a convenient excuse to complain about the "unemployables". I think they have expected too much, too soon, from our educational institutions. However, this does not mean we can ignore the fact that this change is going on all around us. The university must realise that the present value of a degree in purporting to prove that a person is familiar with a discrete body of knowledge, must change within the coming decade. Emphasis will shift to proving that a student has the capacity to adapt to the changing nature of a particular job, and is flexible enough to accommodate new ideas and methods even though they may entirely replace the present method.

Rigid, closely defined assessment procedures do just the reverse. Instead of encouraging a flexible attitude they encourage the inflexible. Questions such as "do you have the right answer?" should be replaced by ones such as " on the available evidence, what appears to be the most suitable answer at this time?"

Democracy in education is important as a means of preparation for involvement in society. The world about us grows more and more complex each day and the rules, which we previously considered inflexible, alter minute by minute. It is important to retain a sense of confidence in our ability to cope with these rapidly changing circumstances. This sense of confidence can be instilled by allowing students to become involved in determining the direction and nature of their education. I believe we are seeing the first signs of such a demand. One thing often forgotten by academics is that the nature and structure of the student population is changing quite rapidly

towards a more even age distribution. Mature age students , probably already some years in the workforce and entrusted with numerous responsibilities, must find it doubly difficult to return to an institution where they are judged to be incapable of deciding how and what they wish to learn. Involvement in their education can also help to reduce an ever-increasing problem on the large campus: alienation. This can become an inspiration to be involved even further, as confidence grows.

Finally, I would like to talk about a concept which I call 'the privileged irresponsibility of academia'. There has long been a view that knowledge was in itself a useful thing, and scientific discoveries were to be made for their own sake. This view has changed, and only in quite recent times.

The increasingly intimate connections between science and industry, science and governmental agencies, and science and military technology, have made it more and more difficult to maintain a clear cut distinction between knowledge and the uses made of it. With this realisation, there has come a call for increased control of what we learn, and for increased education about the implications of what we learn. The time has gone when others were capable of judging this for us.

The production of skilled workers is useless if the factory in which they are to be housed is in the process of falling down about their ears. Our future progress depends on our present involvement, and I leave you with that thought.

QUESTIONS FROM THE FLOOR

Question:

You said that there was too much assessment of our students and then that you felt continuous assessment was good. What do you mean by continuous assessment?

Ms Coulton:

I do believe continuous assessment is a good thing. But basically what I'm getting at is that students should be able to decide, along with staff, what form of assessment they want. Formal assessment means, to me, that the staff or somebody else is projecting down what the student should be studying. Continuous assessment can involve the students themselves in what they want to study. There is room for it.

Question:

Would you suggest that assessment procedures should be decided by the majority of students in a body or by individuals?

Ms Coulton:

I think it should be up to interested students to decide and the group should agree to the form of assessment they want. It comes out of the group itself, I don't believe there is any room for individual types of assessment. That's a personal point of view, though.

SECTION II

PANEL DISCUSSION

RESPONSE BY DISCUSSANT, DR. JOHN KLEINIG

I gather that my position is to raise problems. Because I am rather more philosophically inclined, the sorts of problems that I raise will be at a fairly theoretical level, although they have some important practical implications. Any serious learning situation must involve assessment, in the sense of some feedback concerning the extent to which learning has taken place within that particular situation. That assessment may take a variety of forms, and this morning we have been using a number of terms sometimes interchangeably, sometimes inconsistently, relating to these different forms which assessment may take. So I want to distinguish fairly quickly these different forms of assessment, in order to clarify one aspect of the discussion.

Assessment may take the form of an evaluation, in which some evaluative notion such as good or poor, weak or excellent, or something of this sort, is used with respect to the learning that has taken place. Here an appeal is made to standards believed implicit in what is being evaluated. Or the assessment may take the form of an appraisal which is a form of evaluation having reference to factors which are external to what is being assessed. If I am simply judging a person's work as good or poor, then I am speaking of it in terms of criteria which are internal to the kind of work it is, but if I am appraising someone's work then the evaluation generally has regard to some external purpose. So I appraise you work with respect to some job for which you have applied, or scholarship, or something of that sort.

As well as those two notions, there are notions of ordering and ranking, which are comparative notions in respect of which people are related to each other. In the case of ranking, some kind of hierarchy of ordering is involved. There are other terms that we commonly use, and were used this morning, the terms grading and marking. When we are grading people we are generally making a holistic judgement of their work. That is we are not judging one specific aspect of that work

e. g. their ability to apply their material, we are making some very general, overall judgement of the work. It may be comparative, it may not be. Most of us in University-type situations are caught between two stools here. There is some pressure on us to make grading comparative, as well as absolute, simply because there is an expectation that there should be, say, 10% A's, 15% B's and 35% C's or something of that sort, and if our grades fall too far outside those guidelines we may be asked to defend them.

As well as this, we use the terms exam and test - exams a form of assessment usually taking place under fairly artifical and closely supervised conditions at the end of a course, and often contrasted with some form of continuous assessment the latter being rather more informal and taking place as the course progresses. Tests may form part of that continuous assessment.

Now I make these distinctions, not simply for the purpose of making distinctions, but because they show something important about our purposes. Were our purposes in the Universities or C.A.E. strictly educational ones then, although we would still need to be involved in assessment, I doubt if we would need to be involved in most of those particular forms of assessment that I have mentioned. There would be no need for exams or grading or marking or so on. Now, of course, education is not what the University and C.A.E. is for, and I think it is important that we recognise that. The University and C.A.E. are public facilities and whether or not we like it or not their purposes are only secondarily if at all educational. Their purposes are primarily occupational, and for that reason the kinds of assessments that we have to engage in tend to be of the comparative grading-type. If we were seriously concerned about the educational value of University work then we would, for example, be much happier with co-operative work than we are at present; but in fact, we generally insist on grading students individually, and co operative ventures in which they might

wish to engage are seen as problematic This was the case at Macquarie last year, when students in one course wished to be assessed on co-operatively produced work. The Senate was reluctant to accept the results, for, it asked, how could you tell what each person had contributed to this co-operative venture? The things we have to recognise like it or not (and I don't) is that Universities and C.A.E.'s are not primarily educational institutions, they are public facilities. And it is for those reasons alone that we are involved in the valid forms of assessment that I mentioned - appraisal, ranking, grading, exams, continuous assessment, and so on.

Now when we come to the question of formal assessment, what are we trying to find out? At least two of the earlier speakers raised this very important question. We could be concerned to determine the student's ability to apply that material, the ability to provide some kind of analysis of a particular problem or to synthesize pieces of information from a variety of areas, the ability to criticise the material etc. That's one dimension within which we could assess the student's work. There are other dimensions, too. We have to think in terms of a student's suitability for some future occupation. If it's Medicine or Architecture, the student's suitability for the professions of Medicine and Architecture needs to be assessed. The same goes for a number of other occupational areas. Now one of the problems thereby introduced into the situation is that there is a great number of things that you may want to assess students on, some of which you will be able to assess formally, but others of which you won't be able to assess very easily at all. Dr. Bandaranayka mentioned, for example, the importance of attitudes in Medicine, and I agree with that. The problem is: how do you assess people's attitudes? You can't do this very easily by any of the formal ways, and because you can't do it formally, it tends to be ignored althogether or very nearly so. As well as that, some of the different purposes for which we have formal assessment may be incompatible with each other; for example, what makes a person suitable for the profession may not sit easily with certain things that one might want to test. A person's creativity may not be suited to current demands of the profession, and if you are examining for these two things you might find yourself in some kind of difficulty.

If we wish to assess a variety of abilities, skills, attitudes etc. we will obviously need a great number of different techniques. This was noted by Dr. Bandaranayka The problem with this is that we work in a situation which does not allow a great deal of time and energy to be devoted to the requisite number or kind of techniques. No doubt, if you read numbers of books on education measurement, you can gradually pick up a whole range of techniques that have been tried in various places and these will expand one's repertoire. But for the most part assessment appears to be restricted to verbal expressions on paper. To some extent this may reflect my own background in Humanities-type subjects, for it's obviously somewhat different in the case of clinical situations or in the case of Architectural Design, where something rather more imaginative, from my side of things anyway can be engaged in. (However I know of a Lecturer in Ancient History at Macquarie University, who has students construct, if they wish, models of ancient musical instruments or other artefacts as part of their assessment). For the most part, however we find ourselves restricted to a narrow range of types of assessment, and we must be, unfortunately, simply because of the constraints of time, numbers, and so on. The time that is required for judging design presentations is frightening, and as Professor Daniels mentioned, the staff cannot really do it the justice it deserves. I think we would run into this difficulty no matter what the subject were we to take the purposes of assessment really seriously.

Two questions arising for formal assessment and raised by Dr. Bandaranayake concern their validity and reliability. In talking about the validity of assessments, we refer to their ability to measure what they are supposed to measure. In the case of reliability, we are talking superficially about marker consistency, but even more deeply than that, about whether marks really represent a fair assessment of the person's work. With respect to validity, we have very great difficulties, I think that the terminal examination situation that Dr. Bandaranayake favours is highly problematic in this regard because people are here working under very artificial conditions. In Medicine moreover they are also working under very highly competitive conditions. There is more at stake than simply

passing an exam, there is a future at stake. As well as this, exams take place at a particular point of time, and at any point of time some people are in positions of temporary stress. It might be influenza, menstruation, distress in connection with a relationship it might be a number of other things, any of which may substantially affect a person's performance. A more difficult problem, as far as the validity of such assessments is concerned, concerns the commensurability of the different factors which we are trying to assess, analytical skills versus critical skills, the amount of knowledge gained versus ability to apply the information and so on. How much weight do you give each of these different factors, and how do you then relate them to each other? I am sure this is one of the factors which makes examiner reliability or unreliability such a notorious thing. Different examiners have in mind different, but quite legitimate, criteria. They may also weight them differently. But even if they have the same criteria in mind, they cannot keep them all in mind at the one time. It is moreover quite an art to construct questions that test just what you want them to test and nothing else. Few of us have developed the art. I mention one other problem about terminal assessments, and this is the problem of feedback and defence. Apart from the rather juridical overtones, I like very much the idea of the jury system, because it does offer some opportunity for explanation, for defence, for some feedback, for some interchange between the people involved. This important aspect of the learning situation does not generally operate in terminal assessment. Even worse, when people receive only a grade, they get no information concerning their weaknesses and strengths. I think that in any form of assessment we should do our utmost to allow for some kind of feedback and dialogue. Again, I speak more from the point of view of someone working in the Humanities, where knowledge so-called isn't something fixed, simply to be imparted and regurgitated, at some particular point in time. The whole area is full of contentious matters, and, even though staff ought to know more, I don't think we should presume that their assessments are unquestionable. There is a great difference between say X knows more than Y and saying that X's assessments are final and unchallengeable. I think we often confuse the two: there ought always to be some mechanism whereby students can engage in dialogue with those to whom the university gives responsibility for providing a final assessment.

QUESTIONS FROM THE FLOOR

Kevin Harris, School of Education:

At a conference on Assessment at this University three years ago people seemed very concerned about the mechanics of assessment without actually getting to what I felt was the crucial issue. You have raised this crucial issue by recognising that the University functions in two ways, as an educational institution and as a public facility. On one hand we assess students' work against either absolute or normative standards. On the other we assess it in terms of public utility. That sort of distinction was blurred by all three earlier speakers.

The first speaker, talking about the jury system, was able to say that what was being done could be equated to University words such as 'distinction'. That sort of equation is far too easy and overlooks the problem. Unfortunately having raised that crucial issue, you have then abdicated and said that we just have to recognise that the University is a public facility.

The conceptual distinction of recognising the University as a public facility is in respect of accreditation. The University is accrediting students, so assessment and accreditation get tied very closely together. On one hand we assess and say ' you are doing very well, this is marvellous'. On the other hand we accredit, which means we say 'you passed our standards', 'you have a B.A. or an M.D.', or 'you can now be a doctor', and so on.

I think the crucial issue is whether the university ought to be an accrediting institution. Two questions come out of that. First, the question should it be an accrediting body? You have abdicated and simply accepted that it is going to be one and that's that. Here it seems to me is a huge political question, perhaps one we can't go into properly here.

The second question is whether the University is in fact an accrediting body or whether it simply hides some of the things it does behind this notion of accreditation.

Looking at it we find that in some areas it is accreditative and in others it is not at all. In some of the former the accreditation is in fact false or mystifying.

I'm not sure just what the University accredits people with when it gives them an Arts Degree or when they do General Studies or so on. When a student does History I, II or III, we are not accrediting the person to become a future History teacher. But in twenty years time the Education Department may look at that person's results and say that because they have History I, II and III they can be one.

Now here we have a total distinction between the employer's accreditation and the University's degree, the University accreditation or assessment. I don't think it is absolutely necessary at all for the University to be accreditative. I don't know whether Professor Daniels will agree with me here but I don't see why it is necessary for a student to go through this University or any University in order to become an Architect. A lot of University accreditation may be, in fact, unnecessary ; unrelated or only tenuously related to employment requirements. In any case it is always possible for employment requirements to be accredited by other bodies than the University.

Professor Daniels:

I think I may best be able to answer the last point. In a good many professions there will be an Institute or other statutory body whose members will require of you certain academic qualifications, among other things, before you can practice.

You may not call yourself an Architect until you are registered as such under the Architects' Act. This Act has not many teeth but one thing it does is to prevent unregistered persons calling themselves Architects. To be registered you must have a qualification from a recognised school, have a certain amount of practical experience and have passed an examination concerned with professional experience (mostly office management, contractural administration and so on).

On the other hand you may become attached to or associated with the Royal Australian Institute of Architects; a learned body as well as one that looks at and attempts to control the qualifications of recognised schools, deciding whether they will remain recognised or not. Both they and the Board of Architects visit Schools of Architecture, usually every five years, for this purpose.

Thus there is some pressure being brought upon Universities, C.A.E.'s and Institutes of Technology which have recognised courses to provide an education in the content and attitudes appropriate to the profession. Within this School of Architecture, however there is also a great effort made to coerce and push those bodies into directions of Architectural Education which the School thinks are right. So there is rapport, if you like, between the community at large and the educational bodies on that score.

Certainly in this School of Architecture we would want among other things to be recognised by the Board and Institute of Architects. That would not mean we keep our courses static. We would change and modify them from time to time in respect of the sorts of services architects ought to be offering. Consequently we have to be pushing and guiding the profession in the direction in which we think Architectural Education should go.

Fortunately that rapport is a good one in this case. We talk together on very amicable terms. Certainly the Faculty wouldn't be very impressed by arguments that the content of courses or the teaching method or any other aspect of Architectural Education ought to be modified to satisfy the requirements of these outside bodies. Academics tend to want to stand as independently as they can. I'll leave my colleagues to address themselves to other points in the question.

Dr. Bandaranayake

I think you were trying to draw a dichotomy between the University and the Employer; what the University wants and what the Employer wants. Is that right? This is particularly true when the University is the employer, when it selects people for academic positions. Very often they go on that person's academic record in the University, on how that person has fared in examinations. But he is being employed in a teaching position and his capabilities in teaching are not assessed at all. Now this is becoming increasingly recognised in Medical education. As you, too, may have noticed there is all over the world this new speciality of Medical Education coming up; people being trained specifically in teaching skills and things like that, for employment in Universities.

Dr. Kleinig

I prescinded from the question of whether the University should be a public facility or more concerned with educational matters simply because I don't see a change, in my lifetime, from the University being what it is, which is basically a public facility. On the other hand I think it is very important to be aware that it is one thing rather than the other. Perhaps the fact that it isn't basically concerned with education may constitute a form of criticism we can level at the way courses are organised and to some extent constrained from time to time. I agree in some cases accreditation is rather closely connected

to gaining a piece of paper, presumably so in professions like Architecture, Law and Medicine where there is close liaison between an outside professional body and the University.

But in other cases, for example if you want to be a journalist or a Public Servant or something like that, the University is not an accrediting body in quite the same way. We do tend to find however that employers use University qualifications as a base accreditation and then go on to look for other things. So I guess that if that piece of paper wasn't there, employers would be very worried. So I think, perhaps sadly, there is still some connection between a University Degree and accreditation in this case.

Mr. Hotchkiss:

Our studies are taken primarily by two groups, Commerce and Arts students. I am thinking in terms of Dixie Coulton's comments this morning and trying to look at it from the student's point of view. I observe that, until recently when there has been a shift because of the labour market situation, by and large people coming through Arts are much less concerned about the vocational aspects of their studies than those from Commerce. Numbers of Arts people suddenly wake up in 2nd or 3rd Year and want to change what they are doing, realising they are not well prepared for obtaining a position outside. They say they wish they had the sorts of qualifications the Commerce people have, coming from a more rigid stream.

Perhaps there is a dual function, of education on the one hand and accreditation on the other and perhaps there is a demand from the community generally and from the students for both sorts of things. We should not make the mistake of trying to give one function away in order to overdo the other. I suspect strongly that if we took the sort of approach advocated by Dixie Coulton this morning, to introduce much more student determination of what ought to be taught into the strongly vocational and accreditational streams, we would in fact sink ourselves in that direction.

On the other hand I strongly support the view that from a purely educational perspective there does exist in practice much greater freedom to introduce what students want, but we should keep these two things separate. Students have a choice of which stream to go for and we should leave it that way.

Ms Coulton :

Could I comment on that? Here it comes to the question of what view of society you have, what role the University should play in this whole thing, and what sorts of value you place on being well prepared. You have a particular view of what being well prepared for society is. You mentioned Arts students as less well prepared than Commerce. What exactly is this 'well prepared' thing you are talking about?

Mr. Hotchkiss:

Simply prepared for a job. That is in fact the way students present it to me. Last long vacation I examined an arrangement we have to give students some experience of what it's like in the Industrial Relations field by placing people unpaid, inside trade unions for a period. We placed people on a paying basis with employers in various Personnel and Industrial Relations areas. Following that a couple of my returning students (one Commerce-Law and one Arts-Law) who had been working together said they would like to talk over their experiences during the previous long vacation. The Arts-Law student regretted very greatly that he lacked some of the skills the Commerce-Law student had acquired as part of the degree requirements but which the particular Arts by-laws prevented him from acquiring. They are thinking in terms of a career. That's what they are asking for.

Ms Coulton:

I think what you should look at here is the question of who should be dictating the terms. Should the employers be dictating to the

Universities and determining which areas we should be studying in? Should they be saying ' Listen - this what we want, this what we expect'? I would really disagree with that in a lot of cases.

If you are encouraging people to think, to evaluate, to adapt to the changing economic structures, you are going to produce people a lot more confident, competent within themselves to be able to adapt to these changes. I think you are taking the point of view that employers are out there and the University's job is just to turn out skilled products to fit into that system. If you follow that sort of thing the people who make the decision in our society will be the employers and not the Universities.

I think our reason for having this conference is to assess what sort of direction this society should be taking and how we should possibly be changing industry and other institutions to the benefit of all. You see it comes down to two things. Whose point of view do you take, and who dictates the terms?

Mr. Hotchkiss:

Our particular Department has a reputation with employers of presenting them with people who are not in fact sufficiently sympathetic to their points of view. However there exists a minimum set of skills, as distinct from attitudes, that will enable these people to have any effective chance of ordinary things like getting a job.

Ms Coulton:

I'm not denying that there is of course a minimum set of skills required in any sort of employment. But you will have a lot more responsible, competent, thinking and more positive sorts of persons in our society if you start overcoming the alienation which people face all their lives by giving them an opportunity at University to decide the direction of their studies and how it fits into the society itself.

38

Professor Daniels:

If I could make just one comment here that may show that Architecture is sympathetic to the view just put forward, in the sense that it sees its students as essentially vocationally oriented. New courses, begun this year, offer two opportunities. One is to take a Bachelor of Architecture course over five years, to graduate and be equipped to practice Architecture. The other is a three year Bachelor of Science in Architecture for those who want to study Architecture but aren't vocationally oriented as far as they see the community's demands at the moment.

That course has an Honours opportunity and, having gained it, the opportunity of doing Ph.D. and further study. We have responded here to a great need in regard to accreditation. The number fronting up for the latter course may be small because it is not yet well known. Or perhaps there is not a great demand for it because most people coming to a university are in fact vocationally oriented. But one thing we have done, even in this vocationally oriented Faculty is to recognise that the University is to be an educational facility as well as just a public facility. I think that is a very important role for it to play. I believe that if a University ceases to do that the whole body of knowledge, the whole educational process and the whole way in which the community might go forward would indeed begin to descend rather than to ascend.

Mr. Burford: Chemical Technology

It's a bit of a shame there are not more students here. The small number of undergraduates here goes against the best interests of the staff. Students have the Alternative Handbook in which assessment procedures come up each year and which I think is very valuable. I'll address my comments mainly to Dixie because I think she has set out to be a little more controversial and has pushed a fairly standard sort of student line which I must admit I mainly agree with.

I think students should share assessment procedures. You find out through examination committees that some terrible things go on. A different issue, but one that does come up, is whether students are getting worked harder. I hear that all the time. Certainly in Chemical Technology the number of hours of lectures and quantity of material hasn't materially changed in the last ten to fifteen years, but students are more aware of their rights than they were fifteen years ago. Perhaps students in those days were worked harder and took it without complaining. Now, perhaps, students are game enough to go against parental repression and prepared to stick up for themselves in a University situation.

I think you said that in our society these days formal assessment is not important any more. Now there is a distinction between formal and informal assessment about which you should be very careful. Today people are being assessed all the time, whether on attitude or skills or whatever. For instance status is still, as it has always been, very much a concern of people today -rightly or wrongly. I think our society as it is presently set up is always going to require some form of assessment, be it a number, a distinction, a grade a bit of paper. Society prefers numbers to feelings or opinions and that won't change no matter how you do the assessing. Societies take a long time to change.

On a minor point you mentioned mature age students. The latest I've read is that they are the best performing group; in fact the best group identified was the housewife group. That seems to contradict your statements.

Lastly, when students are speaking up for their rights and complaining more than ever (quite often rightly so-- they should have complained years ago) one of society's responses is to regard them as 'spoiled brats'. More and more students are coming out with fewer of the conventional skills society used to demand. I think the public has

to be educated to see that in fact we have developed better as human beings even though we may not have the professional development as was understood in the past.

Ms Coulton:-

To clarify the point about mature age students, I didn't at all deny that they were getting the best marks in the University but I did make the point that formal assessment, in which the staff are preaching down to students is more of an insult to those people. They have actually worked around for a while and become competent in quite a few skills. Then they come here and are told ' this is how you are going to work'.

You were saying that the work load hasn't increased but that people are more aware of their rights at University. I think that the alternative handbooks and information we get through our Students' Union Education Department shows that the actual workload has increased in a lot of subjects. But putting that aside, I agree that people are becoming more aware and there is room for students to have control over how much they want to learn and what they want to learn. This should be a right in itself.

There has been very little evidence of staff letting people be involved in assessment. As a result people don't know what it is all about. They would find being involved difficult at first because they have never been given a choice as to what they want to study, right through the system . Never. Even in the workforce people are continually being told what they should do. I disagree with the grading system because under it people are working for extrinsic rather than intrinsic rewards. All through the school system you work for external marks instead of being involved in a subject because you want to.

In the work situation you work for money rather than because you like the job. You said you don't think these things can change; it is a long, hard process. I agree but the point is that we

shouldn't say that and then stop. I think that, to a student, University is a place where there is time and access to relevant information and for them to be socially responsible they must play some role and start changing the work situation.

You said people are being assessed all the time. Yes their control over their work is continually being taken away from them. We have to push for people to have more control over their work and you start within the education system. Students must have more control over what they study. This will then lead to changes in the rest of society.

Mr. Burford:

What I mean is that society as a whole is slow to respond to that sort of change. In the meantime you have to embark on some sort of public relations exercise else it will appear you are promoting something intrinsically wrong.

Ms Coulton:

I would agree with you there and I think that is a role the University can play. It can be socially responsible and plant these ideas within the society itself. If you do give students control over whether they do or do not want assessment you are going to produce a far more competent, whatever that means, and more confident person at the end. They will start to see the whole of society in its proper perspective and get a general understanding of what is going on. They will then understand a lot of the contradictions within the society simply because they are not being narrowed down to a particular little field where they can see X and Y and Z; they can only see A.

I think it is a fundamental role of this University to make people aware of how everything is structured and how they fit into it and to be able to understand it. Unless we start to do this at University we are going to get a very sick society indeed.

Questioner:

I am very interested in Group Assessment. I used it this year and last but ran into some problems with student reactions. They like to work in groups but hate getting a group mark.

Professor Daniels:

We have done this in Architecture from time to time but we have always found it very difficult to award grades. We have occasional excursions, which we recognise have great educational value, and most students have three or four such opportunities during their course to be involved in group work. They talk to one another and become involved in the problem more deeply than if they were by themselves. In Architectural Design the product that comes out of group work is more like the camel than the horse. The solution is a consensus of a group, consequently tending to be less innovative than otherwise. The results from group work may thus not necessarily be very good but it is an important process because of the great gains students make in the way that solution is evolved.

Each student gets exactly the same grade. Of course, we are often aware that students A, B and C have really pulled their weight, but D has gone along with the tide. He gets a little bonus, if it really is a bonus. But , if we take our previously described attitude towards University education, that bonus doesn't last for long once he gets out into the workforce or the community at large. He is quickly recognised for what he is.

Dr. Kleinig:

It seems to me very important that people learn to work with each other. The problem when you do encourage this in the University is that people don't know how to. They have been working on their own for so long that in a joint effort they can't communicate with each other to set it all up and develop it. Often the result is a pretty unfortunate pastiche of ideas. I think this is a disaster.

Of course assessing people as a group carries with it all the general problems of individual assessment, but it is crucially important to encourage working together and by the time they get to us they have lost much of the ability to do that. They may get encouraged in Primary school but by the time they go to Secondary school it is a different ball-game.

Ms Coulton:

I agree it is a lot more difficult to encourage group learning and carry out group marking. That should begin in Primary school and the whole grading system should be abolished there at the Primary then at the Secondary school. If we take your attitude it would mean we could never have any sort of group learning or group involvement at all.

Dr. Kleinig:

That wasn't my point. I do encourage it. The sad fact is that it is often a flop.

Ms Coulton:

A lot of these ideas are new to the University, in fact new to people right through their education, but you have to start some somewhere. Why indeed do you need to have individual assessment of any form when you have people learning together? We have had group work in different subjects in the Arts School. People worked together and the lazy person who doesn't act responsibily is reprimanded by the others. Why should it always be up to a person 'up there' telling people they are lazy and reprimanding them? Isn't it better if people work together and each is criticised by the others in the group ?

That will come simply because within a group they are allotted a certain task and if they don't do it they will be criticised by the others; a far more positive thing than an employer, teacher, staff or whatever reprimanding them.

SECTION III

STUDY GROUP REPORTS

INTRODUCTORY NOTE

University teachers and students have few opportunities to meet colleagues from other Schools and Departments (or even their own) in a setting where, free from the formal constraints of teaching and learning, they can converse about issues of common concern.

The Study Groups provided an opportunity for participants to discuss such issues of concern in greater depth than a panel discussion would allow.

The subsequent exploration of assessment issues was rather loosely structured and not intended to be comprehensive. It served to reflect the particular interests and concerns of those present. The reports which follow are edited versions of records made by participants who volunteered to be reporters and we are grateful for their work.

The editor apologises for liberties taken with the original report drafts and takes responsibility for any errors, omissions or misinterpretations.

STUDY GROUP PARTICIPANTS

GRADING & ASSESSMENT

F. Ashe
E. Baker
E. Bryan
W. Kearsley
V. Mack
I. Salasoo
D. Smith
T. Torda
C. Webb
C. Wieneke

CONTINUOUS ASSESSMENT

M. Augee
S. Barnes
R. Bierford
A. Hall
R. Zakarevicius
D. Meredith
M. Pearson
G. Rimmer
S. Saszczak
P. South
S. Tirmazi

ASSESSMENT AS FEEDBACK TO STUDENTS AND STAFF

T. Austin
M. Bailey
E. Barrett
R. Byron
R. Ford
J. Kennedy
V. Paul
A. Sinclair

INVOLVING STUDENTS IN ASSESSMENT PROCEDURES

D. Boud
J. Bruck
P. Cancian
C. Crowley
W. Holmes
S. Prokhovnik
L. Savage
I. Sefton

ASSESSMENT OF CLASS PARTICIPATION

S. Ambirajan
M. Armstrong
M. Crouch
W. Hotchkiss
M. Middleton
J. Nelson
R. O'Leary
L. Owens
I. Qureshi
P. Walsh

GRADING ≠ ASSESSMENT

Leader : Chris Wieneke , T.E.R.C..

DEFINITIONS

In an attempt to distinguish the terms grading and assessment and consider their separate implications, grading was defined as ranking people in order, while assessment was felt to be concerned with the performances of persons relative to some specified criteria. Grading is a possible but not necessary outcome of assessment.

PURPOSES

There are at least three purposes of assessment and different ways of achieving each of them. Those considered by the group were :

(a) testing the amount of knowledge acquired

(b) providing a source of feedback to students

(c) providing a source of data for accreditation

The methods of achieving them will vary with the purposes for which the assessment is being carried out. It is important to realise, however, that assessment always involves making value judgements and as such is subject to error.

THE PROBLEM OF PREDICTION

It was discussed at length whether assessment and grading can have predictive validity and it was argued that assessment gives information only about present performance. It cannot be used to predict future performance.

Hence the responsibility for assessing a person's competence in a

work situation belongs with the employer. Teachers should set what they think are reasonable, realistic and objective assessment standards that are related to possible future employment. Certainly we should be equipping students to face a future environment. But does this mean we are accrediting the course or the students?

Perhaps teachers ought to be expected to provide much more than merely a merit grading of students because it has been shown that there is no necessary correlation between grading and later employment performance.

Against this can be argued that there are many fields where academic success is in fact a predictor of later performance. On balance it may appear possible to predict with some reliability who will be the top performers, but we can be much less certain about those performing in the middle or lower ranges.

Is it in any case the responsibility of the assessor to predict later performance? It does not appear as if the university can realistically pretend to be able to turn out fully qualified professionals in any field. All it can say is that a student has passed, done well or so on. It provides a basis for professional skills and it can teach techniques on how to cope with change. But to presume to evaluate a person's potential is potentially dangerous.

THE PROBLEM OF COMPOSITE MARKS

There appear to be real dangers in the common procedure of adding separate marks together to form a composite grade. It is suggested that such marks ought always to be kept separate for purposes of feedback. There are inevitably problems in deciding the relative weightings of different pieces of work to reflect their separate importances in the overall course.

FEEDBACK MECHANISMS

Grading is certainly one possible feedback mechanism for students but there are surely many other ways of doing this. For example examination papers can be given back to students, with the teacher going through the questions indicating to students where they went wrong. There is potentially a wealthof very useful feedback for students from the close analysis of multiple choice examination papers.*

GRADING

If teachers don't give students a grade will the students be disappointed? If so grading may be working for the wrong reasons. Grading simply implies a comparison of persons. But surely it should be equally possible to grade people 'against themselves'? That is against a specified set of criteria, rather than against other students. Such a practice would obviously not prevent us from also ranking them in comparison with one another if we had good reason for wanting to do it.

PURPOSES OF GRADING

Among the several purposes identified were the functions of providing information to students, teachers, the administration and prospective employers. Of course employers can equally well misuse grades as use them properly because a grade by itself is purporting to tell us something about a student, though exactly what it tells is open to different interpretations.

---

* *See Workshop Report on "Diagnostic and other Feedback based on Examination Analysis" elsewhere in this proceedings.*

ASSESSMENT

Assessment, on the other hand, doesn't have to be at all like that. Ideally assessment should provide more total, holistic statements about performance than the very narrow information provided by grades.

CONTINUOUS ASSESSMENT

Leader : Pat South, N.S.W. Education Department

The group highlighted some issues associated with the use of continuous assessment.

THE PROBLEM OF DEFINITION.

Continuous implies assessment that is spread over the whole learning period rather than kept to the end of a course. It is hence to be contrasted with terminal assessment.

The term seems to carry no necessary implications for formality or informality, so either approach or even a mixture of formal and informal would be compatible with the concept of continuous assessment.

GRADING *

It is clearly not necessary that all work contributing to continuous assessment be graded. At one extreme in a small highly motivated Honours group informal assessment mainly for feedback to students would be the norm and this need involve no grading whatever. In

* *See also the Study Group Report "Grading ≠ Assessment" elsewhere in this proceedings.*

many other cases, however, some degree of continuous assessment involving grading is a very useful counter to the much maligned practice of making a final examination the sole assessment criterion. Credit can be given, in a formal manner, to virtually all work done progressively during a course if this is felt desirable.

This is not without its weakness, however. Earlier work during a course may for most students be lower quality work done while they are getting to grips with the subject. The feedback function of assessment at this stage is of unquestioned importance so that students can learn by their mistakes. For all work to be graded count towards the final course grade may well inhibit risk-taking and even stunt the educational progress of some. Combining course-work grades with those from a final examination to some degree smooths out this possible inequity but considerable room for debate seems to remain as to the optimal approach to take.

FEEDBACK

The possibility and indeed probability of students getting learning feedback is enhanced by the use of continuous assessment and forms a major argument in its favour. * A second advantage is the motivational effect of putting continuous (though perhaps gentle) pressure on students all through a course and avoiding the worst aspects of examination cramming.

A third advantage is its versatility since multiple assessment events would seem to encourage the use of a variety of modes of assessment, unlike a single examination which tends to suggest a single mode. Further it may enhance the likelihood of teacher and students paying more attention to process within a course and less attention to end-product.

---

* *This is examined in greater detail in a Study Group Report "Assessment as Feedback to Students & Staff" elsewhere in this proceedings.*

But a major issue often arises, in practice, in relation to the feedback function. If class numbers are large (or even moderate) how can a teacher solve the practical problem of actually generating feedback information for each student after each test or assignment? If time is the problem the use of multiple choice tests may be the remedy?* Another is to provide answer sheets after a test so that students can develop their own feedback insights into where and perhaps why they went wrong.

However if essay or assignment writing is an assessment mode the time consumed in reading and providing meaningful written feedback criticism can be daunting in anything other than a very small class. Certainly the use of tutors as assessors can help though this raises reliability problems over and above those already known to be associated with essay assessment. The one-to-one tutor/student relationship may be an unattainable ideal in the classes most of us face.

STUDENT WORKLOAD

Not all the problems with continuous assessment are experienced directly by the teacher. An issue that can dominate staff-student dialogue is that of the tendency for student workloads to increase under continuous assessment regimes.** It is easy for teachers to provide too much work in a subject to satisfy themselves that they are teaching the subject well. If all such work is graded the burden on students can be very great. Under a constant barrage of grades the pressure can become intolerable, the student facing assignments in one subject or another week after week with very little time for giving thought to anything else. We would do well to ask ourselves if this were an educational 'advantage' that continuous assessment provides?

---

* *See a Workshop Report on "Diagnostic and other Feedback Reports based on Exam Analysis", elsewhere in this Proceedings.*

** *See Staff-Student debate on this point in the Panel Discussion earlier in this Proceedings.*

We must look critically at the effects of any assessment system on both parties - the students and the teachers. Continuous assessment is certainly no exception to this rule. It may well be an approach that requires us to make even greater efforts at clarifying and comprehending assessment effects than those methods it claims to supersede.

ASSESSMENT AS FEEDBACK TO STUDENTS AND STAFF

Leader : Eve Barrett , T.E.R.C.

The group examined the possible feedback value of each form of assessment commonly in use and explored the evident benefits to both student and teacher brought about by access to feedback information.

Optimal availability of such feedback is probably experienced when we employ a mastery learning approach.* The large class sizes we generally face make it unlikely that we will routinely adopt this method.**

The longstanding and generally unchallenged University practice of retaining final examination papers came under fire from the group. The practice can be criticised for denying students any opportunity to become aware of basic misconceptions or incorrect ideas they may still carry with them, and similarly denying them the chance to

---

* *See the report of a workshop in criterion-referenced assessment elsewhere in this Proceedings.*

** *Keller Plan approaches to individualised learning as well as some forms of student contracting can be used in large groups.*

remedy these before entering professional life or even the next year of their present course.

In addition, this robs the successful student of considerable opportunity for gaining a sense of achievement from seeing the work that was assessed as good or excellent. Further it prevents any possible dialogue between student and teacher which, if it occurred, might let them identify variant interpretations of essay or problem-type questions. Similarly if papers were returned students could often spot arithmetic errors in calculating sub and grand-totals and have a rational basis, which they are at present denied, for requesting a recount.

If on the other hand we used assessment alternatives such as frequent short, non-graded tests (anonymous or otherwise), larger mid-session tests, marked weekly assignments or regular tutorial exercises, the relative weightings given to the separate modes of assessment can seriously influence the significance the student places on both the assessment event and the feedback received from it.

If we employ a tutorial system we ought not to neglect the opportunity this gives whereby lecturers can gain assessment feedback about themselves and the course. It is often quite a simple matter to schedule regular lecturer-tutor meetings specifically to discuss student responses to course material.

The correct answer or the worked solution to each problem should wherever possible be made available to students as soon as feasible after a test is completed. In the case of objective tests, this can of course appear to compromise the principle of developing a bank of valid unseen questions from which to construct future tests*. It remains, however, that highly intelligent and perceptive students

---

** See however, the Workshop Report on "Feedback from objective and other forms of examination", elsewhere in this Proceedings, for ways by which this can be achieved without compromise of security.*

can be disadvantaged in multiple choice tests through wrong interpretation of questions when they see deeper or more subtle meanings to the statement of the question than the examiner intended. Unless papers are given back there is no possibility of this being discovered and justice done. Nor is there of the questions being reworded for future use since the real problem of divergent interpretations remains unrecognised.

If a course is of the more highly structured type, in which assessment and subsequent feedback events occur in a tightly scheduled and predictable sequence, it is possible to argue that student independence may be under threat. Certainly it would appear that the scope for exercising student initiative must be less in such a case. So there is clearly room for continued debate on this more general issue of the relationship between course structure, assessment and feedback procedures, and optimal levels of student autonomy in courses.

INVOLVING STUDENTS IN ASSESSMENT PROCEDURES.

Leader : Dave Boud T.E.R.C.

The group examined the twin issues of whether students should be involved in assessment procedures, and if so how this could be achieved.

On the first point, overall academic opinion would probably favour student involvement, though actual involvement at present is much greater in Arts-based than in Science-based courses. This may reflect a difference between the two cultures, inasmuch as scientific emphasis on the preservation and extension of knowledge through using particular experimental and conceptual tools may lead to an underemphasis on the social implications of the knowledge thereby gained. Certainly the scientific disciplines emphasise the

value of substantive knowledge and cognitive process to a high degree. It may then appear that only a person who 'has' the knowledge can properly assess a person who is still learning it. Yet in arts and humanities participatory learning seems to be easier to emphasise and the resultant dialogue between students and teacher creates an environment in which it is relatively easy to let students become more involved in their own assessment.

To the extent to which these generalisations are valid an implication is that for students to be more involved in assessment procedures in science-based courses would require a quite fundamental shift in our view of the nature of scientific knowledge and its transmission, with less emphasis on the virtues of possessing highly developed abstractions and more on the personal response of the student to the experience of being introduced to that knowledge.

A likely result of increased student involvement in assessment procedures would be increased motivation. Students can help assess themselves and an examination of how this could be done formed the second task of the group. At Hawkesbury Agricultural College, Richmond, the School of Management and Human Resources operates a course titled "Graduate Diploma in Extension" designed for extension workers, adult educators and community training or development leaders. It is run as a participative workshop with no formal lecturing programme. The staff role is to guide the learning process rather than to lecture and students help define their own learning aims and develop the structure of their own learning projects. They consequently assess themselves as part of the organisation of the course. The processes involved are emphasised throughout and students assess both themselves and their peers under the guidance of one of the teachers.

Hence students are learning behaviours that are entirely congruent with those they will expect to practice in their careers in the outside community. This is most appropriate in the case of such careers as those of extension workers. Could it be applied in, say, large first year basic science courses - either in part or in whole - at the organisation level? If it were, students would identify more readily with subject goals by being able to identify the goals themselves. Externally imposed goals are rarely understood; indeed there is often a mental set directed against the goals, so they become rejected.

Perhaps it would be feasible to design a carefully planned programme of induction to the subject matter in which students identify their own need for the subject. That would mean a response, on their part, in the affective domain. This initial attitudinal set would need to be reinforced and students could proceed to select their own subject goals from those available in the area and then work toward them under teacher guidance. Inevitably groups would identify similar sets of goals and this would ease the overall administrative problems.

These sorts of analysis and proposals are clearly exploratory. They probably deserve, however, to have their possibilities seriously examined by those best qualified to do so. This may hopefully lead to increasing student involvement in assessment procedures in a much wider spectrum of disciplines than occurs at present.

ASSESSMENT OF CLASS PARTICIPATION

Leader : Merilyn Armstrong, T.E.R.C.

Out of a range of different opinions as to the meaning of class participation, a majority view emerged that the group should consider it to be students' oral participation in class discussion. Class discussion exists to offer students, among other things, an

active learning role and opportunity to develop oral skills that could have future career relevance. The group expressed unanimous belief in the worthwhileness of class discussion as a teaching method in principle but feared that, in face of the evident difficulties experienced in making discussions work, there are grave dangers that they may degenerate into unstructured lectures.

One way to encourage oral participation is to attempt to assess it. It would be possible to offer a percentage of final marks to class participation and some teachers operate along these lines. Those objecting would doubtless claim that successful participation levels derive more appropriately from quite different factors, such as the atmosphere within the group and the manner of the teacher's response to the students ;factors intrinsic, rather than extrinsic, to the discussion method.

Studies appear to show that indeed the assessment of class participation (A.C.P.) can encourage student contributions and active preparation though this is challenged as being not always of a desirable type. The motive of highly participant students can be that of gaining marks and in addition the spontaneity of discussion can be severely inhibited.

Assessment can actually disadvantage some students, notably the shy ones and those for whom English is a second language. Moreover there are practical difficulties facing a teacher who wants to assess class participation. First the need to decide which aspects of participation should be assessed, the quality or quantity or both? Then to decide how to assess the chosen objectives? Then other factors such as whether or not to give separate marks for attendance and for non-oral participation? These are all difficult decisions and concensus of opinion on any one of them is very unlikely in any group of teachers beyond perhaps a general agreement that, among these factors, quality could hardly be ignored in any valid attempt to assess student contributions to a discussion.

In discussing these problems most members of the group felt that A.C.P. was nevertheless worthwhile, the major point in its favour being that it does evidently tend to encourage students to talk. Moreover if we adopt an 'assessment package' point of view and value variety in our assessment procedures in the interests of accommodating better to individual student differences, then unquestionably A.C.P. is another alternative to add to our package. Some students are better at communicating their knowledge in writing while others prefer the oral idiom, so in order to be fair we ought not to bias our assessment in favour of one group only. It could even be argued that as oral communication is by some measures the most important general vehicle of social communication this alone ought to sway our argument in its favour. One way or the other there seems a very good case for encouraging and even assessing oral skills among our students.

REFERENCES

1. *ARMSTRONG, M.T. Assessing Students' Participation in Class Discussion.* *Assessment in Higher Education* *3, 186-202 (1978)*

2. *CHISOLM, R. Assessing Class Performance. In CONNOR, D. & WARNEKE, C. (Eds.)* *The Assessment of Students*, *Resource Document No. 3. Tertiary Education Research Centre, U.N.S.W., Sydney, (1975).*

3. *HOPKINS, W. Assessing Students' Behaviour in Group Discussion.* *Research in Education* *5, 35-51 (1971).*

SECTION IV

WORKSHOP REPORTS

INTRODUCTORY NOTE

It was realised in planning this Conference that there were current issues, of both theoretical and practical interest, on which participants may wish to concentrate their attention for an extended time. A programme of workshops was designed to allow groups to meet for up to three hours under the leadership of a person skilled in or closely familiar with the topic. Workshop leaders were invited from both this University and several other local Institutions.

Workshop styles varied greatly, as the reports indicate. On one hand were those that were content-oriented, in which the leader presented a substantive body of informative material. On the other were the process-oriented where the leader planned the workshop around the on-the-spot contributions of the participants.

Reports have been written by the leaders themselves and, as published here, have in most cases had only minimal editing.

Several documents being either background articles or analyses relevant to several workshops have been published as appendixes. Where there existed longer published articles or books of relevance to a workshop, footnote reference has been made to them.

The select bibliography that completes this volume comprises, with a few additions, the body of materials collected by T.E.R.C. and made available for display and consultation at the Conference. Though not exhaustive it is probably representative of the enormous range of recently published work, both local and overseas in origin, on Assessment in Higher Education.

DIAGNOSTIC AND OTHER FEEDBACK BASED ON EXAMINATION ANALYSIS

Leader: Dr. Max Dunstan, T.E.R.C.

Each year, students and staff of the University spend a considerable amount of time and effort on examinations. It seems rather important that the wealth of information which is known to be obtainable through the analysis of examination results, should be distributed in such a way and at such a time as will be of maximum benefit to all those interested; students, lecturers and administrators.

The purpose of this workshop was to introduce participants to the already existing avenues by which this information retrieval and distribution can occur. The role of examination results in evaluation of the learning process has, it is felt, been too long ignored. Examinations can play a vital role in diagnosing strengths and weaknesses in students' learning. This is true especially, though not exclusively, where multiple choice tests are set, whether at the end of a course or during it.

Computer facilities at this University (housed at the Educational Testing Centre) enable detailed analyses of the frequencies of good and poor students making particular errors or holding particular erroneous beliefs to be communicated to lecturers in tabulated form. In addition, the frequency of the correct response to each question can be provided.

It is possible, however, to arrange for overall results to be returned to students as a general appraisal of their learning and, in some courses where appropriate, students' papers can be returned to them and individual weaknesses noted.

These already-operating techniques enable students to gain very great study benefits without the security of the test questions being compromised - an important matter of concern in view of how difficult good multiple choice questions are to write. The information available

by this means includes;

1. A letter to each student advising

   (a) the exam mark in relation to others who took the test, and

   (b) the performance on the test in terms of topics plus

   (c) appropriate references for remedial work.

2. Teaching-learning diagnostic information describing the strengths and weaknesses of the whole class in terms of subject topics which have been provided by the lecturer.

3. Evaluation of the examination itself, by means of the identification of defects and suggestions for rememdying them.

It is also possible, in a reasonably comprehensive examination, to offer information to students on their success in terms of educational skill (as distinct from course content). This requires particular collaboration of the lecturer, and results in valuable guides to students that facilitate learning improvement.

The workshop studied these procedures and others through the structured programme of topics listed below:

1. Introduction
   - feedback and its place in learning
   - the usefulness of feedback
   - kinds of feedback
   - evaluation of feedback

2. Feedback to lecturers and students, from multiple choice exams.

3. Construction of multiple choice examinations for diagnostic purposes.

5. Feedback services available for problem-type and oral exams based on analytic rating of categories of questions.

   . services to students
   . services to lecturers

6. Evaluation of lectures and courses
   Construction of standard instruments and individualised questionnaires.

7. General research and survey services

ASSESSMENT OF LABORATORY WORK

Leader: Dr. Ian Sefton, Sydney University.

PURPOSES OF ASSESSMENT

There first exists the problem of whether assessment of actual laboratory work (rather than its final outcomes) is necessary at all. This is highlighted by the questions : (1) what should we do about a student who has not done any laboratory work but performs brilliantly in a formal exam which includes questions related to laboratory? (2) In any case, how can we achieve an appropriate balance between grading of laboratory work and other kinds of assessment not related to the laboratory? We recognise nevertheless that graded assessments given during a laboratory course may provide feedback and motivation, especially to grade-conscious students.

GOALS AND CRITERIA

It is often difficult to define the real goals of laboratory work and to separate the assessment of defined objectives from the actual laboratory work which, in many cases, is merely a learning activity

directed towards the final goals. Specific questions arising from this include:-

1. In marking reports, what weighting (if any) should be given to written expression, neatness, presentation, organisation of data and interpretation of data?

2. How can presentation be separated from content ?

3. Should we assess students' attitudes and, if so, how?

4. How much weight should we give to manipulative skills?

SOURCES OF ERROR

We have already touched on some problems concerned with validity and reliability of assessments. Others include:-

1. In large courses how do we ensure comparability amongst grades awarded by different markers?

2. How do we assess individual work or achievement in cases where students work in groups, particularly when some students may be dominant?

3. Even if students do individual work, what can we do about the problem of copied reports or access to collections of previously marked reports on the same topics?

4. Can we prepare a sufficiently large collection of meaningful experiments or exercises for assessment purposes?

5. Can such a collection of exercises be standardised?

6. What do we do about cooking of results?

7. What are the meanings of grades when pressures from colleagues, students and institutions force us to mark in the range, say, 50% to 100%? How can such grades be compared with those obtained in formal exams?

8, How can we avoid the influence of extraneous factors (e. g. the personality of the student) in making judgements?

OTHER ISSUES

1. In many courses, particularly those with large enrolments, it is difficult to find sufficient staff time for proper supervision and thorough marking of reports.

2. As a consequence, students do not get immediate feedback.

3. There may also be a need to train staff in the required assessment procedures.

4. There is a conflict between a demonstrator's roles as teacher and assessor. Students may be inhibited from seeking help if they think it may affect their grades.

SOLUTIONS

Some ideas emerge from a discussion of the above problems. Perhaps we should:-

1. Not assess laboratory work at all

2. Defer assessment until a single final, written exam.

3. For specific skills, use mastery tests, from which the result is a clear yes or no

or

5. Impose some kind of normalisation on grades from different markers.

or

6. Give markers sufficient information about the overall pattern of grades awarded so that they will carry out their own normalisation.

7. Use peer-group assessment.

ASSESSMENT IN PRE-SERVICE TEACHER EDUCATION PROGRAMMES

Leaders : Dr. Lee Owens and David Smith, Sydney University

The workshop perceived a very real need for a "portfolio" of assessment data that a graduating teacher-education student can use in seeking employment.

By using various forms of assessment, teacher education institutions gather considerable information about the students who have completed courses. There are at least three uses for such information:

1. assessment as diagnosis, i.e. feedback of evaluations and judgements to the student with the intention of promoting further personal and professional growth.

2. assessment as a decision about employability, i.e. qualitative information upon which an employer can base a decision to hire or not to hire

3. assessment as placement information, i.e., having made a prior decision to hire, an employer may be interested in matching a particular position with a uniquely suited individual.

Two touchy questions arise immediately:

1. Who is to provide the information to the employer?

   (a) the college or university administration?

   (b) the student himself?

   (c) the teacher education advisory officer (in the case of the State Education Department)?

2. What information is to be provided?

   (a) anything the employer asks for?

   (b) what the students request to be sent?

   (c) what university or college staff, in their wisdom, choose to release?

In general terms, a constructive sequence of events could occur as follows:

(a) discussion among teacher education institutional representatives, teacher education advisory officers, representatives of the main employers, and students about what information is germane and what form it might generally take

↓

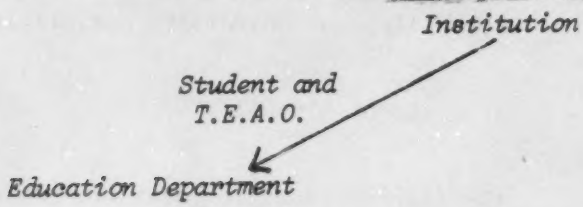
(b) transmitting to the student the necessary and/or requested specific information so that the student can prepare a "portfolio" which represents who he is, and what he has accomplished, and how he is suited to a teaching job.

↓

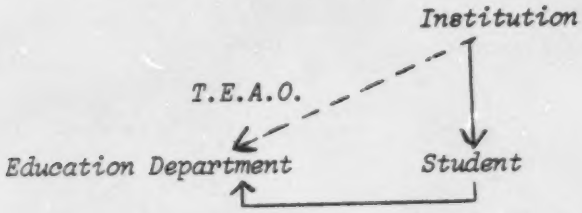
(c) negotiations between employer and candidate for a teaching job during which the "portfolio" information can be discussed.

(d) appointment to a job

The process is not as straightforward as this, because at the moment the major employer (for example) receives its information from the student via the Teacher Education Advisory Officer:



It is proposed by the workshop that this be amended in the following way:



The student therefore would have an immediate need to learn how to accumulate the necessary information, and how to set it out effectively in the form of a 'portfolio'.

A three part "portfolio" is suggested, the first two parts being base-line data that might be suggested to each student to prepare, and an optional expansion of base-line information:

Part 1

A brief, standard application for employment form (such as the Education Department T633).

Part 2.

A curriculum vitae containing the academic record, practice teaching report(s), organisation memberships, special interests/hobbies, involvement in community affairs, publications.

At least two problems can be foreseen:

1. is the practice teaching report to be a final summative document, or a series of reports from each individual teaching episode?

2. is the curriculum vitae to be only a set of guidelines, issued to the student for him to do his own thing, or a printed form that is standard for an institution, or a set of pages suitable for a loose-leaf presentation, to be included or deleted as appropriate?

Part 3

A supplementary package that might include letters of recommendation from master teacher(s), sample teaching units prepared in courses or in schools, copies of publications, recommendations or supportive testimony from others such as employers or members of voluntary organizations who can comment on suitability , or special experience that is readily adaptable to extra curricular activities within a school program.

ESSAY MARKING

Leader : Stan Ross, School of Law - University of New South Wales

The workshop agreed that the term essay should be used on this occasion to include reports, examination scripts, fieldwork assignments, etc as well as the traditional-type essay.

APPROACHES TO ESSAY MARKING

There is probably no other area of assessment on which such a wide variety of views and practices exist. Plagiarism is one example of an issue every essay marker faces, where we each tend to use our own definition of what constitutes the crime and our own rules on what to

do when we think it has been committed. One approach is to regard as plagiarism only instances where virtually the entire essay has been 'lifted' from another source; another is to object only when the writer fails to acknowledge direct quotations from the other source.

Responses to plagiarism range from one extreme, deep intolerance that calls for severe penalties, to the other which finds it a tolerable device on the part of undergraduates who after all, have to get their material from somewhere.

Attitudes and practices concerning low level marking are similarly diverse. This covers the issue of awarding penalties for errors of inadequacies in spelling, grammar, typing, citations, footnotes and bibliographies Again, some regard typing errors as most serious. Others are so glad to receive a typed essay that, though they may correct the errors, they would not subtract marks on that account. Further, some allocate marks for adequacy of footnotes, citations and bibliographies while others regard these as irrelevant issues in a marking scheme. One approach to handling this is to produce either independently or in collaboration with colleagues, a set of guidelines for students on how to write essays. This would explicitly state the marker's expectations. Some who have tried it report that even this may not be adequate to meet students' needs when they come to the actual writing task.

ESSAY STRUCTURE

(i) Introductions: opinions diverge on the fundamental matter of what precisely constitutes an introduction. If lecturers cannot agree, what hope is there for students to know exactly what is expected by the way of introduction to an essay? Some are evidently capable of writing up to three pages of irrelevant information before starting to answer the question. Such lengthy introductions clearly confuse the whole concept of word limits. There is even a wide divergence of attitudes as to how strict one ought to be about word limits. One constructive view is that many of these mechanical aspects of writing essays involve the acquisition of skills that can readily be explained and taught to students. Were we to teach these skills, we would

be justified in regarding them with appropriate importance when marking the essay.

(ii) Content: Most of the marks will be awarded to this, which is generally the whole point of the essay. In perhaps most cases understanding of the subject matter would be ranked highest in lecturers' priorities of what to look for when marking. But this hardly solves all our problems. We still have to decide what to do with, say, the 'photocopy' student who makes a pastiche of an enormous amount of relevant material drawn from many sources but whom one suspects doesn't really understand the subject at all. Whatever approach we take to this problem it may well be advisable for markers to have at very least, a clear marking scheme which distinguishes between thought components of the content (e.g.knowledge, analysis, application, presentation structure, logic, etc.) which will presumably be given most of the mark-weighting and polish components (such as references, spelling, etc.) which are given a much lesser weighting.

MARKER EXPERTISE

How important is it for markers themselves to be expert in the actual subject matter of the essay? This can be an especially tricky problem when students are given a range of topics to choose from or where they select their own research topics. Few markers will feel themselves equally at home in every possible field in which they have to assess students' work. Is it irresponsible, nevertheless, for a teacher to set essays on topics in which markers are known to be inexpert? It is common for tutors to be given guidelines in essay marking, telling them the main things to look for and to credit. In addition it is quite feasible for markers to consult together beforehand and discuss the subject matter. Done with care, this can lead inexpert teachers to learn much from the experience of marking a set of essays on a topic somewhere outside their own field. It remains possible that marking standards would tend to vary between expert and inexpert markers, some being more severe and others more lenient.

However there exist techniques such as cross-checking and double-marking, particularly in borderline cases or possible failures,

GROUP MARKS

What are the marking problems that specifically arise when students are allowed to produce work in groups? The main problem seems to be that of deciding whether to award a common group mark to each participant, or to award separate marks on the grounds that some have contributed less to the final product. Individual testing, perhaps orally, could give evidence to support the latter action but would not this be contrary to the whole principle of group work? There does appear to be good justification for common group marks on the grounds that students themselves can best exert pressure on unproductive members. It would, however, seem necessary to distinguish between instances where on the one hand a single 'product' is produced by the combined action of the group and on the other where the group collaborates and shares ideas and resources but separate members finally put together their own products to submit for marking.

Collaboration can of course occur even under ordinary conditions of individual work through informal exchange of ideas among friends. It is important to consider how justified we would be in reading sinister implications into this practice. It may be that the isolated or individualistic student is at a genuine disadvantage here through not discussing ideas with others. There is an analogy with laboratory work, in which it is quite usual for groups to work together but submit individual reports But what do we do if two almost identical essays are received? Which one gets marked down? The second one to hand in the work or both students?

FEEDBACK

The more comprehensive we try to make feedback to students, the longer they will probably have to wait for it. A conflict between these two desirable goals - adequate feedback; and feedback supplied as soon as possible - seems inevitable. If examination papers are involved it may be most profitable to go over the questions in class immediately after

completion of the exam, even though the papers have not been given back. This helps markers , as well, by clarifying what to look for during marking and by sensitising them to particular problems students may have faced, such as ambiguous interpretation of questions.

There is a student objection that challenges the whole point of giving feedback about questions once the marks have already been decided. This problem may be lessened if assessment is both by essays and a final examination, as students may then be more motivated to take notice of written comments and come and talk to teachers about their essay work There is some doubt, however, about the value of markers putting things such as lines or question marks against incorrect or doubtful points. Comments should perhaps be more constructive than this and the general comment at the end of the work should begin by pointing out something positive about the student's effort.

A solution rejected by the group was the idea of producing model answers to show students how they ought to have written the work. This would compromise the whole principle that is implicit in setting essay questions, namely that there are innumerable ways in which to write an answer, all of them potentially good. It follows that there can be no single correct answer. If there were one we would perhaps be wise not to try to use the essay as the mode for its communication.

HOW TO ORGANISE A COURSE WITHOUT GRADING

Leader : Dr. Dave Boud, T.E.R.C.

There has been much debate throughout the world on the appropriateness of grading as a part of student assessment. One of the practical questions which arises from this debate is: if grading were abolished, how would students be assessed? The workshop was designed to specifically address this question. Rather than adopt any particular position in the debate on grading the workshop examined the practical implications of the assumption that grading be not allowed in a course.

To achieve this, four headings were proposed:-

1. The educational purposes of student assessment
2. Information required for these purposes
3. Examples of non-graded assessment procedures
4. Applications of the previous discussion to our own situations.

PURPOSES OF ASSESSMENT

The following were identified

- to motivate
- to establish a standard
- to certify competence
- to facilitate learning
- to provide a pacing mechanism
- to provide feedback of a positive kind
- to establish limits to learning, and provide direction
- to reduce competition and foster cooperation
- to promote critical attitudes

WHAT IS REQUIRED?

If these purposes are to be pursued, then the following is required of assessment

- it should be qualitative rather than quantitative
- it should indicate deviation from established standards derived from the discipline, the teacher and the student. These should be indicated appropriately depending on the level of the student.
- 'outputs' (that is, the students' work) must be in some visible form

- overt criteria for judgement
- detailed, specific comment
- evidence of group activity
- observation of the process of learning, not just outputs

- awareness, by the student and tutor, of the criteria to be applied.
- caring, sympathetic feedback
- knowledge of the aspirations and needs both of the students by staff, and of the staff by students

EXAMPLES

The following were discussed. Many of them did lead to grading because of institutional requirements, but they all have considerable educational merit without the grading function.*

a) Self - and peer - assessment in a cooperatively planned course, including use of a self-assessment schedule.

b) Use of Mastery Learning approaches in, for example, courses run using a Keller format

c) Contract learning in which criteria for successful completion of work is negotiated in advance between students and staff

d) No formal assessments in W.E.A. Adult Education classes.

e) Use of assessment 'profiles' these being descriptive accounts of student performance.

APPLICATIONS TO SPECIFIC SITUATIONS

No record was kept of the specific applications to courses run by the participants although the general problem of the interface between educationally desirable assessment procedures and institutional requirements was discussed. Nor was any attempt made to reach general conclusions or to recommend activities to others.

* *There is insufficient space to give an account of each, but further information and references are available from the workshop leader.*

SELECTING APPROPRIATE ASSESSMENT METHODS FOR A COURSE

Leader : Jackie Lublin, N.S.W.I.T.

The workshop formed small groups of participants to engage in their own problem identification. From this it emerged that there is current concern about:

(i) the relationship between teaching and assessment

(ii) the time consumed by assessment procedures

(iii) the problem of relating assessment to course objectives and

(iv) the choice of methods appropriate to assessment in specific learning situations (e.g. group work, laboratory work, large groups).

The teaching-learning situation is a complex interlocking web from which it is impossible to completely isolate assessment issues. The connection between different types of assessment methods and course objectives was one axis chosen for discussion. No doubt appears to exist that assessment methods are intimately related to course objectives, but other factors such as which purposes assessment is intended to serve would have to be taken into account before making a choice in a specific case.

Whichever assessment approach we choose will have its own effects, inevitably, on student learning. For example the present move to continuous assessment has had its effect in taking from students the responsibility for long term planning and organisation of their own workload.* Similarly the sort of learning that an anticipated project or essay assessment may encourage would differ significantly from the sort demanded by a time-limited objective-type test. There are other issues as well that need consideration when choosing between time-limited examinations and other assessment types. One hears arguments that the experience of being under pressure to perform in

---

* *This issue is discussed further in a Study Group Report on "Continuous Assessment", elsewhere in this proceedings.*

the examination situation gives valuable training for facing similar stress points in ones career. Consistency is necessary, however, so if assignments or reports are employed it follows that time deadlines for these should in fairness be waived only on grounds similar to those for which a supplementary examination would be granted.

The workshop explored the wide range of available assessment modes from the commonest - such as written exams, multiple-choice tests, essays and projects - to the less common types such as peer assessment, self assessment and viva voce tests. A mark given for class participation or attendance can also become part of assessment. On what basis do people make final decisions about students passing or failing?

Educational measurements are in general notoriously imprecise. Moreover when we add different marks together we generally get great bunches of students near the middle of our distribution. It is likely that the so-called formal assessment procedures frequently fail to tell us what we really want to know in order to separate the class into those 'pass and fail' groups or the A-B-C-D classification expected of us. It seems that in practice we generally invoke quite intuitive and informal criteria such as rapport, esteem, past work, interest displayed and so on. These are factors admittedly quite unreliable as well as invalid but we employ them in order to accomplish the job expected of us as assessors of students.

What sort of assessment does one choose, when does one use it and how does one carry it out most appropriately in a particular course? It seems that course objectives form the starting point. From here the teacher derives samples of expected student abilities, particularly in the knowledge and skills domains, which are susceptible to some form of quantification and measurement. Then one makes decisions about timing and about terminal versus continuous assessment, taking into account the particular pros and cons of each in ones own situation. Perhaps a judicious mixture of each is optimal in most cases since students seem to appreciate the reassurance offered by continuous assessment.

Yet in some subjects it is ridiculous to attempt to test the main objectives until a student has been through the whole course and can tie things together. Perhaps the most important objectives are those we test when we are asking students to tie things together and achieve an overview of the whole subject. If so they will deserve the heaviest assessment weighting.

It seems from the experience of this workshop that it is probably impossible, and maybe undesirable as well, to attempt to isolate a limited topic such as this from the wider field of issues in teaching and learning. It certainly appears that there are no easy or indisputably correct answers to the vast majority of questions practising teachers raise in this area.

CRITERION - REFERENCED ASSESSMENT

Leader: Michael Bailey, Macquarie University

This report describes the use of paper-and-pencil and paper-and calculator measures in employing mastery (or Criterion- Referenced) tests in teaching. For readers unfamiliar with this field, a more detailed description of Criterion-Referenced Measurement appears as an Appendix later in this proceedings.

THE PASS MARK

An approximate solution to the problem of determining the passing mark in a mastery test may be obtained by the rather drastic assumptions involved in Lord's formula. If every item in a mastery test is equivalent to every other item, we may regard the testing process as a sequence of independent binomial trials with a constant probability p of success in each trial. The number of successes for any person is then given by np, where n is the number of items, in theory, and the standard deviation of the number of successes (items right) is

np(1-p), by a well known statistical result. Of course p will differ for each person, and we estimate np by knowing the number of items that person passed. If n is fairly large (say 15 or more), the passing score for the test may be taken to be that value of np which allows n to be within 2 standard deviations of np - in other words we choose the smallest value of np satisfying the inequality $np + 2\sqrt{np(1-p)} \geq n$. In practice this is most easily solved by guessing the number and seeing if it works out; it is usually around 80% of the items.

RELIABILITY

No satisfactory statistic has yet been agreed on as being the analogue of reliability in a mastery test, and none may exist. The papers by Subkoviak and Juynh referenced in the bibliography offer possible approaches but certainly not paper-and-pencil measures.

ITEM ANALYSIS

Several procedures for item analysis have been suggested. The earlier ones are well reviewed in a paper by K. D. Crehan*. Most of these are of dubious value, but the modified Brennan index of discrimination may have some application. This is a generalisation of the index of discrimination for norm-referenced tests and is simply the difference between the proportion or percentage of people who passed the item and the test and the proportion or percentage of people who passed the item and failed the test. It is assumed that a "good" item is one which contributes to the classification of masters and non-masters, and thus classifies in the same way as the test does. The larger this index is, the better the item should be.

Two other indices have proved of more value in work done at Macquarie on criterion-referenced test item analysis. They are similarly motivated, but somewhat more specific. If we score 1 for a pass on the item and 1 for a pass on the test, and otherwise zero we can (with a moderately sophisticated pocket calculator) calculate the correlation

* *Item Analysis for Teacher-Made Mastery Tests*, Journal of Educational Measurement 1974, 11, pp 255-262.

coefficient between the item and the test. As both are scored either 1 or 0 the resulting correlation will be the so-called phi coefficient, and this will be the analogue of the biserial coefficient used for item evaluation in norm-referenced tests. It will, however, be influenced by the difficulty level of the item, and this influence can be removed by dividing it by the coefficient phimax, to obtain the adjusted phi-over-phimax correlation.

With a pocket calculator this is most easily calculated as follows: after calculating phi, find the total number of people who have passed the item, and the total who have passed the test. Choose the smaller of these numbers say k. Enter the number pair 1,1 k times, then make up the larger number by entering the number pair 1,0 an appropriate number of times and finally enter the number pair 0,0 sufficient times to make the total number of people who took the test. Thus if p people took the test, q passed the test and r (less than q) passed the item, there will be r entries of (1,1), q-r entries of (1,0) and p-q entries of (0,0). The resulting correlation will be phimax and dividing phi by it will produce the adjusted coefficient. Even with a pocket calculator which computes correlations automatically this is the most awkward statistic to compute of the ones given here (though it is readily programmed for a small computer), but it is apparently the best single approximate statistic measuring item-quality in a criterion-referenced test.

A simpler statistic is more directly related to the actual pass-mark used in the test. If we consider only those people who either just obtained the pass-mark or failed by one mark, and thus those for whom the item may have been of critical importance in determining their success, an index of item-test agreement will be the proportion of those people for whom the item and the test were in agreement, that is those who passed the item and the test or who failed the item and the test. This index is the simplest to calculate and appears to have some value, though it is not as effective as the previous one.

Copyright note: This material is previously unpublished and the copyright in it is hereby reserved by the author. Any reproduction, by any means , is welcomed by the author provided that the source is acknowledged in any further dissemination.

SECTION IV

APPENDIXES

SOME NOTES ON CRITERION-REFERENCED MEASUREMENT

Michael Bailey , Macquarie University

The use of tests to assess skills and capacity is a fairly recent development. It appears to have originated in China in the form of the nationwide examinations in the verbatim knowledge of the literary classics which were used from about the Tang Dynasty period (around 600 A.D.) to select civil servants. The procedure, which was strictly competitive in character and not associated with any system of education, continued to be used during periods of strong central government in China for well over a thousand years without influencing practices in other countries; until well into the nineteeth century measurement practices in Europe were confined to Universities, and there consisted of largely ceremonial disputations between the scholar and his teachers.

During the nineteenth century the growth of the bureaucratic state brought with it two developments which led to a need for the measurement of capacities. The need for administrators, together with a dissatisfaction with the practice of appointment by knowing the right people almost universally accepted until that time, led to the adoption of measurement procedures very similar to those of the Chinese to solve a similar problem. The procedures used would now be classified as norm-referenced assessment; a number of people competed in an examination (often, as in Britain, actually in the knowledge of Classical Literature almost exactly on the Chinese model), and those whose performance was best got the best jobs in the bureaucracy. Shortly thereafter, national education systems developed, and with them came demands for accountability; the taxpayers' money must be shown to be having its due effect in producing an educated workforce. In developing countries such as Britain, France and Germany attempts were made to measure the outcomes of education by finding out how many school pupils knew the alphabet, multiplication table and catechism. In England this movement led to the notorious system of payment by results, one of the earliest uses of what we would now call criterion-

referenced measurement of mastery learning, and it is perhaps in these murky origins that we must find the beginnings of the movement toward criterion-referenced measurement.

Even at that early stage the uses of the two broad categories of measurement practice can be seen to have taken shape. Norm-referenced measurement is most valuable when a relatively large number of people compete for a relatively small number of rewards. It does not have to be closely related to the purpose for which the selection is to be made. Its early uses in Western societies were perhaps philosophically based on the notions of the elderly Plato in the Laws; Golden people (usually the children of Golden parents) being those best fitted to rule. The major advantage of norm-referenced measurement is that it does not require a precise statement of requirements; any kind of test will do, and the process of sorting out who is best can take place when we see what is produced. Thus this sort of measurement process is the only one any rational person would use for judging international piano competitions, awarding medals at the Olympic Games, or choosing ten people from a thousand applicants for scholarship awards.

Criterion-referenced measurement, on the other hand, is far more appropriate when we are interested in whether certain standards have been reached. Its use is limited mainly by the requirements that these standards must be capable of precise definition beforehand. However, we might think that its educational uses would far outnumber those of norm -referenced measurement, because we are constantly needing to know (or at least certainly should be) whether students have acquired skills needed to proceed to the next stage in a learning sequence. With the increasing demand for accountability and the attainment of standards in education, criterion-referenced measurement is likely to increase in importance, because norm-referenced measurement cannot be used for assessing absolute attainment except in a very indirect and awkward manner which reacts adversely on its proper use in competitive situations. Optometrists, driving test examiners, and anyone assessing the attainment of minimal objectives would be foolish not to use criterion-reference measurement techniques. If we start

to investigate the relative usage of norm- and criterion-referenced measurement practices in education, however, we will find that norm-referenced practices are far the more common, and it may be valuable to consider why this should be so.

The prevalence of norm-referenced measurement practices in education may be attributed to a combination of two factors. First, these techniques have the advantage that they avoid raising the question whether teaching and learning have been effective for the students as a whole, by focussing attention on relative rather than absolute achievement. While there is some worry about this, as instanced by the recent statement that H.S.C. examiners will not necessarily award full marks to the best candidate in a subject if they decide they should not, it is hard to see what basis for objective judgement in terms of absolute standards could be obtained at the highest level of attainment. If those examiners have solved this problem, it is to be hoped that they will soon divulge their procedure, because it is precisely at the highest levels of attainment that criterion-referenced measurement has so far proved impossible. Second, norm-referenced techniques can use a powerful statistical methodology. This has been developed very largely for the precise purpose of making comparisons between the relative attainments of different individuals on various measures of intellectual performance. It was never intended as a way of measuring whether certain standards had been reached, but rather was associated with the work of the mental measurement theorists whose interest was either purely theoretical or related to questions of eugenics, this latter concern being still very much with us in the form of the I.Q. controversy. In passing we may note with regret that the results of norm-referenced tests believed to measure relative capacity have found their chief use in the hands of the ignorant to justify limitations on the absolute attainment of certain groups of learners, a use lacking both theoretical and moral justification. Thus we see that the failure to distinguish between the kinds of information provided by norm- and criterion-referenced measurement can have wide-ranging practical consequences; it is not a mere academic quibble. The development of effective criterion-referenced tests has been hindered by the lack of appropriate methods of analysis by which

results can be evaluated. During the 1970's however, techniques have been developed which make effective analysis of criterion-referenced tests and mastery learning sequences possible. These notes conclude with an attempt at a fairly non-technical explanation of the procedures which have been developed, but first some consideration must be given to the question of constructing criterion-referenced tests.

Criterion-referenced measurement has become of increasing practical importance as the use of the mastery learning model has increased, and the procedures involved in the development of criterion-referenced tests closely reflect the requirements of that model. The most controversial aspect of the mastery learning model is the assertion that given enough time anyone can learn anything.

This aspect of the model does not form part of the consideration for the construction of criterion-referenced tests, although it can be allowed for by the related domain-referenced procedure associated chiefly with the name of Hively. Domain-referencing requires the most stringent set of specifications for test construction. A finite domain of possible items which could be used to test the required skill must be precisely defined. Test construction can then be a random process in which items are randomly sampled from the item domain. The difficulty with this attractive procedure lies in the very precise specification needed to implement it. For example suppose we state the objective; "The student will correctly add two one-digit numbers". Empirical studies readily show that this is a very different skill when carrying must be done than when it should not be done. Again, the introduction of zero will disturb the item, the required knowledge being of a rather different kind. Thus an item domain may have to be as precisely defined as including all those one-digit additions which require neither carrying nor the use of zero. It can readily be seen that this kind of precision is neither attainable nor desirable in the majority of educational applications.

Much wider use may be made of the principles of criterion-referenced measurement than of domain-referenced measurement. While we are still

restricted by the need to define our objectives precisely before we can see whether they have been attained, it is probably not necessary to go as far as the more enthusiastic supporters of behavioural objectives have done. If we wish, we may adopt their rather theoretical approach to test construction and define our objectives in terms of actual test performance ("The student will make no mistakes in spelling a list of 20 words randomly chosen from List X"). However the price of adopting this approach is that we neither make allowance for error of measurement nor relate our objective very closely to what is communicable to someone who doesn't speak the jargon. Everyone sometimes misspells a word they normally spell correctly (leaving aside such favourite errors as "miniscule"). Furthermore, and perhaps even more importantly, our interest is not really in Professor X's list of words but in the general universe of words which may be encountered at a certain stage of reading and writing development. Thus it is reasonable to argue that we should seek a formulation of criterion-referenced measurement which allows for some latitude in specifying the performance required and the conditions under which it will be produced. Such a formulation requires a method of handling the concept of error of measurement; various attempts have been made to devise a method and some of them will be found in the references at the end of these notes.

Criterion-referenced assessment has sometimes been taken to mean no more than the precise specification of objectives, often in terms of Bloom's taxonomy. If it is not carried to extremes, this approach may have useful application in the investigation of whether skills in a particular cognitive domain have been developed, and to what level. Certainly the use of precise specifications of basic skills cannot be bad, even in tertiary education. A multiple-choice test in a second-year education course at Macquarie University provides an interesting and perhaps not untypical illustration of poor measurement practice; of 54 items 51 were classifiable at only the knowledge level in Bloom's taxonomy, while the other three did not aspire beyond comprehension. Any attempt to define the areas of competence required and the level of performance to be reached would surely have resulted in some hard thinking, if not actual improvement of the test.

If we do take criterion-referenced measurement to require only the precise specification of behavioural objectives, we do obtain the valuable benefit of infallibility. We know that our measurement procedures are valid because the outcomes we wish to produce are defined in terms of the overt performances of students as we interpret them. To adopt this approach requires greater self-confidence than many of us possess, but the prospect of avoiding all the tedious process of item analysis and test improvement which less self-confident evaluators must undertake is very tempting. The valid use of such measurement practices must be confined, however, to skills at a fairly low cognitive level. A test on, say, the principles of norm-referenced test construction really cannot be so self-evidently valid that mastery of the required (complex) skills is testable by means of a fixed and given set of items without any error of measurement. In particular the required skills can certainly be analysed into components, some of which stand in hierarchical relationship to others in the sense that mastery of simpler skills is required before mastery of more complex skills can be achieved. Learning hierarchies may be broadly classified into three types, which we may call necessary, facilitative and contingent. In facilitative hierarchies, where association between the possession of skills is commonly found, the simpler makes the more complex easier to acquire but is not essential. An example might be the fact that lorry drivers can usually drive cars also. Contingent hierarchies exist in empirical data, but have no apparent reason other than perhaps the general way in which people develop; for example children usually walk before they can write. Necessary hierarchies are of particular interest to us because they offer several ways in which we might attempt to analyse and improve criterion-referenced tests. The reason for this is that if we know A is needed for B and our test assesses a student as passing B but failing A, there is something wrong with our test or our method of determining success or failure on it.

Various attempts to devise methods of analysing criterion-referenced tests have been proposed. In general the simple ones seem rather unrelated to the requirements, as for example the procedure of testing

students at the beginning and end of a course and retaining the items on which most improvement was shown-almost as guaranteed to build the teacher's ego as the norm-referenced procedure where you know that someone will be top and the average mark will be 50 however incompetent you are. The more sophisticated methods have often been based on Bayesian procedures or rather complex formulations in terms of decision theory, and have been directed in the main to assessing the test as a whole and in isolation, without looking at the structure of skills within the area of competence being tested. The methods to be proposed below attempt to avoid both extremes and to offer both complex precise and simple approximative procedures for analysing criterion referenced tests

ANALYSING MASTERY AND SEQUENTIAL STRUCTURES

The assumptions underlying the model of test responses proposed here are rather weaker than those of the traditional trait-theory model associated with norm-referenced measurement .In particular no assumption is made concerning any quantity of capacity, ability or knowledge possessed by a respondent. Each item or skill is regarded as a separate entity, and it is assumed that each item can be dichotomously scored, and each skill categorised as either possessed or not possessed by the respondent. This is a necessary assumption for the application of mastery models, since the outcome desired is itself dichotomous. The requirement is not unnecessarily restrictive as we shall see. It has the important consequence, however, that there is some difficulty in formulating a concept which can take the place of the idea of "true score" in classical test theory. The simplest solution to the difficulty for the purpose of an elementary discussion is probably to adopt a mentalist approach, and this will be done in what follows. Thus we shall assume that it is meaningful to say that a respondent "really knows, but has temporarily forgotten" a particular skill or item of knowledge, and that the respondent "really does not know, but has guessed." These mentalist assumptions can be reformulated in statistical terms if necessary, but the process leads to complex

definitions which are unnecessary for understanding the model proposed. Thus our error theory will hypothesise two types of error, called for our purposes guessing and forgetting, whatever the actual reason for their occurrence may be.

For any test, or set of tests or measurements, the outcome for each respondent can be expressed as a sequence of 1's and 0's when the scoring is dichotomous. Under the model we are considering it will not be meaningful to add the number of items a respondent has answered correctly in order to obtain a total score, because if the items test the same area of skill the possible decisions under the model are as usual, mastery or non-mastery. The outcome for one respondent in a 6 item test may be 010111 and for another it may be 111100. Unless all the items are exactly equivalent (a very implausible assumption) it should be evident that these two performances are not necessarily indicative of the same level or state of skill. Under the assumptions of classical test theory, however, we have no alternative to adding the marks. The Rasch-Wright model for norm-referenced measurement also makes it difficult to do other than add the marks, but under the model proposed here we have no need to do so.

A simple explanation of the analytical techniques that may be used requires a minimal knowledge of probability theory. Suppose that we know or can determine from investigation that the probability of someone who really knows an item forgetting it is f, and that the probability of someone who really doesn't know it guessing it correctly is g. (Remember that this does not necessarily mean that forgetfulness or luck is the actual cause of the assessment error). Another way of formulating this is to say that if we test a large number of people all of whom really know the item, the proportion that gets it wrong will be f. If a test consists of two items (for simplicity) we can write the probability of forgetting the items

as f_1 and f_2 respectively.

It follows that because the probability that a respondent does not forget the first item must be $1-f_1$, we can express the probability of various possible observations (results on the test) thus:-

True pattern	Test result	Probability
11	11	$(1-f_1)\ (1-f_2)$
11	10	$(1-f_1).f_2$
11	01	$f_1.(1-f_2)$
11	00	$f_1.\ f_2$
00	11	$g_1.\ g_2$
00	10	$g_1.\ (1-g_2)$
00	01	$(1-g_1)\ .g_2$
00	00	$g_1\ .\ g_2$

It is evidently possible to extend this table to cases where the test consists of more items, or where we are prepared to consider the possibility that the true pattern or state of knowledge of a respondent may be 10 or 01. In any normal test, of course, the probabilities f and g are fairly small, and thus in a six-item test, if the f- values were of the order of 0.1 the probability of a true pattern 111111 being observed as 000000 would be about 1/1000000. We may note that the model, in common with the models of classical test theory, assumes that the probability of error in any item is independent of whether errors in other items occur or not. Making this assumption, we can develop procedures for estimating the population proportion of each true pattern, and the error probabilities f and g, from a reasonable sized sample. Ideally this should be large, but usable information can be obtained from samples where the number of respondents is about 40 times the number of items used. The size requirement is comparable to that needed for the analysis of norm-referenced tests on the classical model.

Estimates of pattern proportions and error probabilities can be obtained by the use of maximum likelihood techniques. These will not be discussed here, but those interested may refer to Lazarsfeld and Henry (1968), Proctor (1970), Dayton and Macready (1976) or Macready and Dayton (1977) for treatments of the technique used (called Fisher's method of efficient scores) as applied to this and related areas. A valuable by-product is the assignment of any observed pattern of responses to the most probable true pattern, together with an estimate of the likely accuracy of the assignment.

A few examples of the use of the proposed method may be given. First, let us consider the example of a mastery test consisting of 5 items measuring a certain skill. There will be two possible response-patterns of true scores, namely 11111 (mastery) and 00000 (non-mastery) if the mastery model is applicable. We can use the analytical technique mentioned above to estimate the proportions of masters and non-masters in our sample, and the guessing and forgetting probabilities for each item separately. If the estimation algorithm fails to converge, or if, having converged, it produces a very low probability of fit to the proposed model, or if the guessing or forgetting probabilities are implausibly large, we must reject the hypothesis that the test is a mastery test. A single item which does not fit with the others will usually be revealed by its large associated guessing or forgetting probability. It is unlikely that in such a short test three items will test mastery in a domain and the other two be independent; if that does happen the method may give unsatisfactory results and we can use a generalisation of it, or an associated exploratory technique, to investigate the situation further. An item which can be easily guessed will be revealed directly by this method in contrast to the item analysis procedures of classical test theory. These latter methods often leave us guessing as to the reason for an item producing vaguely unsatisfactory item evaluation statistics, because these statistics are not directly related to specific features of the model.

As a second example we may consider a test of six items, the first three testing one skill and the second three another. If the first skill is prerequisite to the second, the true score patterns will be 000000, 111000 and 111111. If, on the other hand, the skills are

independent of each other, a fourth true pattern, 000111, will be needed to represent those people who have mastered the second skill but not the first. We can use a method similar to that of the previous example to estimate the proportions of each score-type in the population, but we may find it impossible to estimate each error rate separately for each item. If this is so, we can constrain the guessing and forgetting probabilities to be the same for every item, a device which does limit the recoverable information about individual items but is not too unrealistic in practice because in a given type of test guessing probabilities at least are usually about equal, being about zero for a constructed-response test and about 0.15 for a typical multiple-choice test. In compensation we obtain a method of comparing the hypotheses that skill 1 is a prerequisite to skill 2 and that the skills are independent by means of a likelihood-ratio chi-square test, as well as the usual test of whether the model fits the data. As usual, we also get a scoring scheme which allows us to determine whether or not a given respondent has mastered either of the skills.

Thirdly, we may generalise the estimation procedure still further. Any set of legitimate responses may be used as the model to be tested. If we have five skills that are believed to form a linear hierarchy (or Guttman simplex), the legitimate response-patterns will be 00000 10000, 11000, 11100, 11110 and 11111. Again we can test the hypothesis that the data fit the model, score the compound test to determine the level reached by a respondent when that is not obvious (clearly if a respondent's score pattern is, say, 11100, it would be absurd to propose any other true pattern that 11100 as accounting for it), and decide if other response-patterns appear to be occurring in the sample of respondents. Hierarchies may not be linear, but no matter what their structure, linear, branching, or looping, the general procedure can test for it. In this application we definitely must constrain all the forgetting probabilities to be equal, and likewise the guessing probabilities, or the solution will be indeterminate and the algorithm will fail to converge. This third application allows us to ask the question whether mastery matters, because once mastery or its lack have been determined we can give each person the appropriate

score 1 or 0, and examine the skill in relation to other skills. If mastery does not allow, or its lack does not prevent, progression to a further stage of learning, then it is not of much educational importance except as a terminal skill. Also we can use the procedure with an ordinary general-purpose test to ask what the passing score should be. Consider a test of mathematical skills on entry to tertiary education. If we know that one skill depends on another (calculus on algebra, say) we can use the procedure to determine the pass-mark needed in the algebra and calculus sections of the test, and thus to decide in a more objective fashion than usual what levels of skills are needed.

The drawback of the methods proposed above is that they require the use of a largeish computer, because maximum-likelihood estimation procedures use fairly large amounts of both space and time to operate. This is not these days perhaps a crippling disadvantage in a tertiary institution, since many of them own or have easy access to a large enough computer for the purpose. However, a simple approximation to the proposed analyses may be useful, and can be performed by anyone with a desk-calculator and a bit of patience. For dichotomous scoring of responses the correlation between any two items, or between an item and the pass/fail outcome of a test, may be calculated by the usual product-moment formula. In this instance, the result is the fourfold point correlation, or phi coefficient. Items believed to be testing the same general objective should be correlated at a reasonable level; in fact this level can be shown to be about the level obtained by treating the phi coefficients as if they were product-moment correlations and using the Bonferroni protection level procedure to determine the critical phi-value for significance. If the 10% level is chosen for the test, the required significance level is $0.1/(k-1)$ where k is the number of items in the test. A poor item then shows up as having nonsignificant correlation with the other items in the test. Some improvement in the procedure can be obtained, at the cost of rather more computation, by using the phi-over-phimax coefficient, the ratio of phi to its maximum possible value given the difficulties of the items, and this improvement is particularly marked in the correlation of success on the item with success on the test. The procedure is very similar to the simple method used for some years in norm-referenced test analysis, and an approximation to the even simpler discrimination index method may be obtained by using a descriptive

discrimination index being simply the difference between the difficulty of the item for those who passed the test and its difficulty for those who failed. (Here "difficulty" is used in the technical sense of "proportion who got the item right").

These simple procedures do not assist us in determining what the criterion mark for a test should be, and there seems no simple way in which this can be done. Perhaps we have nothing better than the rule-of-thumb 80% mark which has been used in many contexts for some years, and which has been shown in several studies using more sophisticated methodology to be roughly of the right order of magnitude. Use of the maximum likelihood estimation procedure for determining a pass mark is thus to be desired wherever the outcome is of considerable importance to the people being tested.

In general it may now be confidently asserted that the methods now available for criterion-referenced test analysis can at least be unashamed of a comparison with analytical methods for norm-referenced tests: the rather greater power of the latter is balanced by the wider applicability of criterion-referenced assessment in the educational context.

BIBLIOGRAPHY

(a) General

Airasian, P.W. and Madaus, G.F. Criterion Referenced Testing, in Crucial Issues in Testing ed. Tyler, R.W. and Wolf, R.M. Berkeley, Cal : McCutchan, 1974, 77-88.

Glaser, R. and Nitko A.J. Measurement in Learning and Instruction, in Educational Measurement (2nd Edition), ed. Thorndike, R.L. American Council on Education, Washington D.C. 1971, 625-670.

Gronlund, N.E. Preparing Criterion-Referenced Tests for Classroom Instruction. New York:Macmillan, 1973

Popham, W.J. Technical Travails of Developing Criterion-Referenced Tests, ERIC ED091421, 1974

(b) Specific in relation to this document

Dayton, C.M. and Macready G.B. A probabilistic model for validation of behavioral hierarchies, Psychometrika 1976, 41, 2, 189-204

Lazarsfeld, P.F. & Henry, N.W., Latent Structure Analysis Boston: Houghton Mifflin, 1968

Macready, G.B. and Dayton, C.M., The use of probabilistic models in the assessment of mastery, J. Educ. Statistics 1977, 2, 2, 99-120

Proctor, C. A probabilistic formulation and statistical analysis of Guttman scaling, Psychometrika 1970, 33, 1, 73-78

(c) Some alternative approaches

Emrick, J.A. An evaluation model for mastery testing J. Educ. Measurement 1971, 8, 4, 321-326

Huynh H., On consistency of decisions in criterion-referenced testing J. Educ. Measurement 1976, 13, 4, 253-264

Meskauskas, J.A., Evaluation models for criterion-referenced testing, Review of Educational Research 1976, 46, 1, 133-148

Subkoviak, M.J., Estimating reliability from a single administration of a mastery test J. Educ. Measurement 1976, 13, 4, 265-276

TEACHING THE WRITING OF ACADEMIC PAPERS : A CONTRIBUTION TO DISCUSSION

Peter Bamford, University of Newcastle

One of the most widely used teaching-learning strategies employed in the University is that of requiring students to write an essay or 'academic paper' of some 1,500-2,000 words on a set topic. This particular form of examining students has a long and respectable pedigree and manifestly legitimate functions; inter alia, it serves: to introduce students to the modes of enquiry and the conceptual tools of a particular discipline; to instruct them as to the nature and function of technical exposition; and to inculcate the necessity for academic rigour in the analysis and synthesis of insights derived from reading and thinking.

Recently the University Counselling Service has become aware that a significant and apparently growing number of students, particularly first year and mature age students, are experiencing substantial difficulties in writing academic papers. These difficulties would appear to be traceable to the lack of certain skills fundamental to the writing of such papers. Among the more common difficulties are the following: an ignorance of the proper function of the introduction, the body, and the conclusion in the essay-form; confusion as to the precise meaning of 'key-words' in the topic, such as 'analyse', 'compare', 'contrast' 'evaluate', 'discuss', and the like; an inability to incorporate material derived from sources into the structure of the paper in the form of valid generalizations with supporting evidence; a failure to distinguish between personal and authoritative opinion, and a lack of awareness of the legitimate functions of both in the presentation of a reasoned argument or exposition such as an academic paper demands.

It would appear from the (admittedly small) sample of students seen in counselling for this particular problem, and from their (admittedly partial!) verbalization of their feelings of confusion and anxiety as to what precisely is expected of them in the presentation of an academic paper, that in this area of teaching and learning there is a legitimate cause for concern for academic staff.

It is probably fair to say that the students presenting with these difficulties in counselling represent the "tip of the ice-berg" in terms of the total number of students who experience confusion and anxiety in varying degrees about expectations held by staff as to the 'correct' principles and procedures to be observed in an academic paper in their particular discipline.

The skills of writing academic papers, like other skills, are teachable. Clearly the mechanical skills of citing references and acknowledging sources are minimal, lower-order skills in this regard and are relatively easily acquired by students by practice over time. The larger, more complex, higher-order skills of presenting a coherent, sequential exposition involving the legitimate uses of evidence and opinion are not so easily acquired. Yet, surprisingly, the lower-order skills are more often more explicitly taught than the higher-order ones. Such higher-order skills, moreover, are arguably not most effectively taught by cryptic or diffuse end-of-paper comments, although the legitimate functions of such comments are not to be denied, if they are explicit enough. Students, apparently, are often taken aback by lecturers' comments (something we've all experienced!) not so much by the legitimacy or otherwise of such comments but because of the discrepancy between their expectations of what they thought the lecturer required of them and what he actually expected, as evidenced by his comments. In other words, students are often unclear as to what criteria of assessment are to be applied to their paper until after the event, and even then, considerable confusion may remain.

It may perhaps be argued, in the case of first year students who have just completed their secondary education, that they will begin tertiary studies already possessed of the requisite skills, both lower-order and higher-order as defined above. It is true that some students coming straight from school to University will already have received a good grounding in logical exposition and the conventions of 'essay' writing; other students, both post-secondary and mature entry will not have received such a grounding. In addition, the University will naturally require a higher level of such skills as students are progressively required to perform academically at a truly tertiary level.

Mature students in particular, (although those who have completed the Open Foundation course will have the rudiments of academic skills), are even less likely to have an adequate grasp of the nature and function of the academic paper per se. The argument sometimes advanced that "The schools should have done it", or, "It is not our job", would appear to be of doubtful validity.

From the considerations indicated above, it would appear that students might benefit from being taught how to write an academic paper early in their course. On this basic premise, the following guidelines are presented for discussion as possible strategies for achieving this end:

1. Students could be given explicit instruction in how to construct an academic paper. Consideration of what such a paper entails in terms of aims, objectives, methods, and outcomes would seem a logical approach to tackling such a teaching task. It would include an examination of such fundamentals as : the precise definitions of key words and phrases (analyse, illustrate, compare and contrast, critically evaluate, and the like) commonly used in essay topics; the nature and function of the major components of a paper (the introduction, the body, and the conclusion); the presentation of valid generalizations with appropriate supporting evidence; and the legitimate uses of personal and of authoritative opinion. The lower-order skills of referencing sources, foot-noting, compiling bibliographies, setting out quotations would be taught, of course, pari passu, with the higher-order skills instanced above.

2. Students could be required to practise writing an academic paper in part and/or in toto. A possible approach here would be for the first paper to be submitted in a course to be considered a 'learning exercise' in writing an academic paper, and either to be not assessed at all or to count a minimal amount (5%? 10%?) towards the final assessment. Assignments might be progressively weighted throughout the course: Essay I might be 10%, Essay II, 15%, Essay III, 25%. An arrangement

of this nature, or a variant of it would go some way towards minimizing the psychological shock suffered from "failing" one's first paper, which may represent a considerable proportion (20% -30% in some cases) of the year's mark.

3. The provision by departments of a model paper in their discipline as an aid to students in preparing and presenting their academic papers. Clearly, there are valid objections to using the strategy of a 'model paper' if such a paper is to be perceived as one to be imitated slavishly. However, there is a good deal to be gained by students, in terms of learning the conventions of specialized writing, as well as in terms of personal security and growth in confidence, in having an 'approved version' of an academic paper available as a basic guide in their initial efforts in this medium at the tertiary level.

4. Criteria for the assessment of an academic paper should be clearly specified before the event. There is some danger that lecturers may make the assumption that students should know what is expected of them; that they should infer what has not been made explicit. It is less than reasonable to expect the student to interpret generalities about academic standards, analytic rigour, and the like, however well-meaning, and, further to accomplish (relatively) unaided the task of translating such generalities into a written exposition. Students might be told, then, that their academic paper will be assessed on the basis of some such set of criteria as the following : breadth and depth of reading, with special emphasis on reference to up-to-date research; clear exposition of insights derived from reading and thinking; critical evaluation of evidence adduced on the topic under consideration; logical structure of the paper; conciseness of expression; adequacy of foot-noting and/or bibliographic referencing.

Clearly, the definition and illustration of the terms used in the criteria would be a sine qua non of their being fulfilled in students' papers. A 'model paper' would be relevant in this regard.

There are considerable advantages to the student to be derived from the adoption of such guidelines for the writing of academic papers. They would serve to: give him the opportunity of acquiring certain fundamental skills at the outset of his University career; focus his learning, thereby making it more efficient; and, not least, minimize anxiety, by giving him the security of knowing precisely the legitimate expectations of academic staff. Different disciplines will have particular requirements and expectations for the presentation of academic papers, it is true, but there is enough commonality of expectation as to standards for the guidelines indicated above to have a significant degree of relevance across disciplines. For some students, a "failed" paper can be the last straw in terms of a touch-and-go decision to drop out of University or soldier on. Not to take account of the potential for good or ill of such a widely-used instrument of assessment as the academic paper is, it is suggested, both inhumane and uneconomic. And continuous assessment, unless wisely used in accordance with the accepted principles of learning, can easily become, for some students at least, a source of continuous stress.

The advantages to academic staff to be derived from the adoption of such guidelines are also evident: focusing students' learning by the use of agreed criteria may contribute significantly to the effectiveness of teaching and learning; marking, though always onerous is made 'easier' if one is tackling a paper which is based on agreed principles and, not least, it is extremely useful to be able to point to a set of criteria in case of disputed assessment!

Relationships between staff and students are most at risk in the delicate area of assessment - after all, assessment is essentially one fallible human being sitting in judgment on the efforts of another fallible human being - and relationships are most easily strained when confusion and ambiguity are evident on either side of the equation. Justice is more easily seen to be done if the lecturer's expectations regarding academic performance are clear and unequivocal from the beginning

(Reprinted, with permission, from the H.E.R.S.U. Bulletin No.9 University of Newcastle, June 1978)

THE PROBLEM OF PLAGIARISM

The extracts that follow are reprinted from the Notes on Legal Research and Writing of the University of New South Wales Law School.

The Law School's rules concerning the use of writings of others represent a consensus achieved after considerable discussions among students and staff. Issued annually to all law students (they) provide the basic principles and procedures concerning plagiarism, inadequate attribution, excessive collaboration etc. They are taken very seriously by the Law School community as a whole.

THE LAW SCHOOL'S RULES ON THE USE OF THE WRITINGS OF OTHERS

In most courses, part of the assessment is based on essays or other written assignments which the student researches for himself. Unfortunately there have been occasional instances of students quoting passages from published material (or other students' work) without citing the source, and passing it off as their own work. The School took the view that it was desirable to make it clear both to students and staff what is regarded as improper, and what measures will be taken when a student appears to have presented written work which makes improper use of other writings. These rules were adopted by Faculty (resolution 75/76) and are set out below. THESE RULES ARE VERY IMPORTANT AND YOU SHOULD READ THEM CAREFULLY. IF IN DOUBT, CONSULT YOUR TEACHER.

A. PRINCIPLES

Essays submitted for assessment should always give due acknowledgment to the sources which have been used.

The basic principles are that you should not attempt to pass off the work of another person as your own, and you should enable a reader to check in the original sources the information and ideas which you have used. It is difficult to be more precise, but the following are examples of breaches of these principles :

(a) Quotation without use of quotation marks. This is a common, and very serious, breach.

(b) Significant paraphrasing - e.g. several sentences or one very important sentence, which in wording are very similar to the source. This applies even if the source is mentioned, unless there is also due acknowledgment of the fact that the source has been paraphrased.

(c) Unacknowledged use of information or ideas, unless such information or ideas are commonplace.

(d) Citing sources (e.g. cases, texts) which you have not read, without acknowledging the 'secondary' source from which knowledge of them has been obtained.

These principles apply to both text and footnotes of sources. They also apply to sources such as teaching materials and to any work by any student (including the student submitting the essay) which has been or will be otherwise submitted for assessment. A student must obtain the prior approval of his teacher if he wishes to submit to that teacher an essay substantially similar to one which has already been, or will be, submitted to another teacher.

Whether in the text or the footnotes, acknowledgment should be sufficiently accurate to enable the source to be located speedily.

If you are unsure whether, or how, to make acknowledgment, consult your teacher.

Failure to observe the principles referred to above may lead to either or both of the following consequences.

B. REDUCTION IN MARKS

Assessment of an essay will not be affected by any belief on the part of the teacher that the student should be "punished" for "dishonesty".

Accordingly, every essay will be assessed "on its merits," leaving aside considerations of dishonesty. However, failure to comply with the principles referred to in Part A constitutes a defect in an essay and therefore is likely to lead to the essay receiving a lower mark "on the merits" than it would otherwise have received. The effect that such failure to comply will have upon assessment will normally depend upon (a) the length of the passages concerned, (b) the significance of those passages to the essay as a whole, and (c) the extent to which acknowledgment (if any) has been made. Thus, assessment is more likely to be affected significantly if the passages concerned form a large part of the essay, or relate to matters which are particularly important to the essay, or if a source has been used frequently but has rarely, or never, been acknowledged.

It is almost certain that, if detected, any failure to observe the principles referred to in Part A will lead to some reduction in marks, even if the passages concerned are short and quite unimportant parts of the essay. The more substantial in length or importance such passages are, the greater the reduction will be. It is conceivable that an assessment of 0% might result.

Having made such reduction, a teacher may give the student the opportunity to submit extra work on the bases that the marks obtained by the essay are to be modified or replaced by the marks obtained by the extra work. There is no general obligation upon teachers to offer the opportunity to do extra work. Teachers will not require extra work for punitive purposes.

If an assessment has been affected by an alleged failure to observe principles referred to in Part A, the teacher, when returning the essay to the student, will indicate the relevant passages and except when the failure is so pronounced that it is not practicable to do so, will give an approximate indication of the marks lost as a result of those passages. Of course, a teacher may discuss the matter with the student before deciding upon his assessment or giving the opportunity to submit extra work.

On all matters of assessment, the student has a right of appeal to the Dean .

ASPECTS OF STUDENT ASSESSMENT BIBLIOGRAPHY

1 *Anderson, A.W. Scaling Assessments. RUUE Bulletin, University of Western Australia, November 1974.*

2. *Armstrong, Merilyn. The Assessment of Class Participation in the University of New South Wales Law School, Tertiary Education Research Centre, Research & Development Paper No. 45. December 1976.*

3. *Assessment in Higher Education, University of Bath, Educational Services Unit. (This Journal is devoted to the Subject of Assessment and is published three times a year. Editor : Dr. Duncan Harris of the above Unit).*

4. *Barrett, Eve M. The Method of Assessment in Mathematics 10.021 1975, Tertiary Education Research Centre, Research & Development Paper No. 45, December 1976.*

5. *Beard, R. Teaching and Learning in Higher Education, U.K. Penguin, 1970.*

6. *Bligh, Donald and others. Teaching Students, Devon, Exeter University Teaching Services, 1975.*

7. *Bloom, Benjamin J. and others, Handbook on Formative and Summative Evaluation of Student Learning, N.Y. McGraw Hill, 1971.*

8. *Cockburn, Barbara & Ross, Alex. Inside Assessment, University of Lancaster, Teaching in Higher Education Series, No. 7, 1977.*

9 *Cockburn, Barbara & Ross, Alex. Essays, University of Lancaster Teaching in Higher Education. Series No. 8, 1977.*

10. *Conference on Assessment - Report - Bundanoon, Australian National University, Office for Research in Academic Methods 1975.*

11. *Connor, D. & Wieneke, Chris. The Assessment of Students, Tertiary Education Research Centre, Research Document No-3, June 1975.*

12. *Cox, Roy. Examinations in Higher Education. Universities Quarterly, June 1967.*

13. *Dunstan, M. The Setting of Essay Examination Questions. Tertiary Education Research Centre, Bulletin No.1 (Reprinted) 1971.*

14. *Dunstan, M. A Reply to Some Criticisms of Objective Tests, Tertiary Education Research Centre, Bulletin No. 5. (Reprint) 1972.*

15. *Dunstan, M. The Problem Solving Process of Tertiary Students on Verbal Material. Tertiary Education Research Centre, Bulletin No. 4 (undated)*

16. *Dunstan, M. Interpretation of Item Analysis Data, Tertiary Education Research Centre, Bulletin No. 5. (Reprint) 1972.*

17. *Dunstan, M. Measuring Complex Skills on A-K Answer Sheets. Tertiary Education Research Centre, Bulletin No. 10 (Reprint) 1972.*

18. *Dunstan, M. A Check List for Preparing the Title Page and Directions for Objective Examinations, Tertiary Education Research Centre, Bulletin No. 11 (Reprint) 1972.*

19. *Dunstan, M. A Guide to the Planning, Writing and Review of Multiple Choice Tests, Tertiary Education Research Centre, Bulletin No. 12 (Reprint) 1972.*

20. *Dunstan M. Letters to Students, Teaching and Learning Diagnostic Information and Evaluation of Examination Reports. Tertiary Education Research Centre, Bulletin No. 13 1972.*

21. Ericksen, S. Grading & Evaluation, Memo to the Faculty No. 46, University of Michigan, 1971

22. Examination Papers, Australian Union of Students, Melbourne, September 1975.

23. Good, Harold M. Interview marking of examination scripts, Assessment in Higher Education 3 (2) 122-138 March 1978.

24. Group Assessment - The Fight Continues, Adelaide, University Students' Union, 1977.

25. Hall, W.C. and Cannon, R. University Teaching, University of Adelaide, A.C. U.E. 1975.

26. Heywood, J. Assessment in Higher Education, U.K. Wiley, 1976.

27. Higher Education Learning Project, Practical Work in Undergraduate Science, London, Nuffield Foundation Heinneman, 1977.

28. Hill, Barry, The double marking of scripts in university examinations in engineering, Assessment in Higher Education 2 (2) Feb. 87-103, 1977,

29. Izard, J.F. Construction and Analysis of Classroom Tests. Victoria, Australian Council for Educational Research, 1977.

30. Just, F.P. The Case for Examinations, Parkville, Australian Council for Educational Standards Monograph, October 1973.

31 Klug, Brian. Pro-Profiles - a Study of the Uses of Grading and Profiles in Academic Assessment, London, National Union of Students, (undated)

32. Lawrence, Susan and Hayes, R.A. Towards Objective Assessment of Class Participation, Tertiary Education Research Centre, Research & Development Paper No. 19, October 1972.

33. Magin, D. Assessment of students in Australian Universities. Public Lecture, I.T.B. Bandung (T.E.R.C. Mimeograph) June 1976.

34. Miller, C.M.L. and Parlett, M. Up to the Mark - a study of the Examination Game, London, Society for Research into Higher Education, 1974.

35. More Words - Workshop Reports from the Second National Trainee Teachers' Conference, Wagga Wagga, April 1976.

36. National Activists Conference on Post-School Education - Papers - Melbourne. Australian Union of Students May 1977.

37. Popham, W. James (Ed.) Evaluation in Education California, L.A. Mc.Cutchan, 1974.

38. Powell, J.P. Higher Education : A Select Bibliography Vol 3 1970-1975, Sydney H.E.R.D.S.A. 1977.

39. Roe, Ernest. Using and Misusing the Material of Teaching and Learning, Canberra. Australian National University, 1975.

40. Rose, C. and Trent, J.W. Student characteristics and the learning process in Travers, R.M. (Ed.) Second Handbook of Research on Teaching, Chicago, Rand McNally, 1973.

41. Seminar on Continuous Assessment - Report, Tertiary Education Research Centre, April 1970.

42. Rowntree, Derek. Assessing Students - How Shall We Know Them? London, Harper and Row, 1977.

43. Ten Brink, Terry D. Evaluation - a Practical Guide for Teachers N.Y. McGraw Hill, 1974.

44. *University of Adelaide Committee on Assessment - Report (Student Edition) Adelaide, Students Association of the University of Adelaide August 1975.*

45. *West, L.H.T. Student assessment, Notes on Higher Education No. 9, Melbourne, Higher Education Advisory and Research Unit, Monash University, February 1975.*

For further references readers should consult Powell, J.P. Higher Education : A Select Bibliography, Vol 3. 1970-1975 and Supplement to Vol.1 Sydney, H.E.R.D.S.A., pp.87-93, 1977.

Dr. W. R. Albury	History & Philosophy of Science
Dr. S. Ambirajan	Economics
Mr. F. Ashe	Mechanical & Industrial Engineering
Dr. M. L. Augee	Zoology
Mr. T. K. Austin	Student, Medicine
Mr. M. Bailey	Education, Macquarie University
Dr. E. Baker	Mechanical & Industrial Engineering
Dr. R. Bandaranayka	Centre for Medical Education Research and Development
Ms S. Barnes	Social Work
Dr. A. K. Bhattacharyya	Mining Engineering
Dr. I. D. Black	Faculty of Arts
Dr. A. M. Bordow	Behavioural Science
Dr. J. Bruck	General Studies
Ms E. Bryan	Student, Arts
Mr. R. Burford	Chemical Technology
Mr. R. A. Byron	Mechanical & Industrial Engineering
Ms P. Cancian	Student, Arts
Dr. J. A. Cartmel	Mechanical & Industrial Engineering
Dr. A. E. Churches	Mechanical & Industrial Engineering
Dr. D. J. Corderoy	Metallurgy
Ms D. Coulton	Students Union
Dr. D. B. Croft	Zoology
Ms M. Crouch	Sociology
Mr. C. J. Crowley	Student, Arts
Ms S. Dail	U.N.S.W. Union
Professor E.C. Daniels	Architecture
Mr. C. Evans	Education
Dr. R. A. Ford	Mechanical & Industrial Engineering
Mr. A. Hall	Kuring-gai C. A.E.
Mr. C. K. Harris	Education
Mr. J. Henderson	U.N.S.W. Students' Union
Assoc.Professor W. Holmes	Electrical Engineering
Mr. W. E. Hotchkiss	Industrial Relations
Ms S. Jacob	Student, Education
Ms L. Jennings	Kuring-gai C.A.E.
Dr. A. H. Kearsley	Surveying
Mr. J.P. Kennedy	Wool & Pastoral Sciences
Mr. K. Kjorrefjord	Mechanical & Industrial Engineering

PARTICIPANTS

Dr. J. Kleinig	History, Philosophy & Politics Macquarie University
Ms J. Lublin	New South Wales Institute of Technology
Dr. D. Meredith	Economics
Ms V. Mack	Kuring-gai C.A.E.
Ms M. Middleton	Librarianship
Mr. J. Nelson	Librarianship
Mr. R. O'Leary	Marketing
Dr. Lee Owens	Education, Sydney University
Dr. A.C. Partridge	Mineral Processing
Ms V. Paul	Mathematics
Assoc.Professor J. Prokhovnik	Mathematics
Dr. I. R. Qureshi	Applied Geology
Mr. P. M. Redmond	Law
Mr. T. Rickard	Engineering Physics, Hawkesbury C.A.E.
Professor W.G. Rimmer	Economics
Dr. S. D. Ross	Law
Dr. I. Salasoo	Chemistry
Ms S. Saszczak	
Mr. L. Savage	Students' Union
Dr. I. Sefton	Physics, Sydney University
Mr. A. N. Sinclair	Wool & Pastoral Sciences
Mr. D. Smith	Education, Sydney University
Ms P. South	Teacher Education Advisory Office N.S.W. Education Department
Professor N.L. Svensson	Mechanical & Industrial Engineering
Ms A. Symes	Teacher Ecucation Advisory Office N.S.W. Education Department
Mr. S.I. Tirmazi	Food Technology
Professor A.R. Toakley	Building
Dr. T. A. Torda	Surgery
Mr. P. Walsh	Architecture
Ms C. Webb	Economics
Dr. R. A. Zakarevicius	Electrical Engineering.